Underspecification in the macroeconomic Arbitrage Pricing Theory (APT) linear factor model and the role of the residual market factor

by

JAN JAKUB SZCZYGIELSKI

Student Number: u04911807

Submitted in partial fulfilment of the requirements for the degree

PHILOSOPHIAE DOCTOR in FINANCIAL MANAGEMENT SCIENCES

in the

FACULTY OF ECONOMIC AND MANAGEMENT SCIENCES

at the

UNIVERSITY OF PRETORIA

Supervisors:

Prof LM Brümmer

Prof HP Wolmarans

August 2018

DECLARATION

1. I understand what plagiarism entails and am aware of the University's policy in this regard.
2. I declare that this assignment is my own, original work; that all sources used/quoted have been indicated and acknowledged by means of a complete reference system.
3. I did not copy and paste any information directly from an electronic source (e.g., a web page, electronic journal article or CD ROM) into this document.
4. I did not make use of another student's previous work and submitted it as my own.
5. This study was not previously submitted for a degree at another university.

Signature	Date

ACKNOWLEDGEMENTS

I would like to thank my late father, Dr Jan Henryk Jozef Szczygielski, for encouraging me to always study further and to pursue an academic career. Furthermore, my sincere appreciation goes to the following:

- My supervisors, Prof Leon Brümmer and Prof Hendrik P. Wolmarans, who provided guidance throughout the writing of this thesis and engaged in stimulating discussions of the subject;

- Prof Henco van Schalkwyk, the Head of Department at the University of Pretoria, for supporting my numerous academic endeavours, which included conference attendances, funding for submission fees, equipment and software.

- The Faculty of Economic and Management Sciences at the University of Pretoria, for giving me the opportunity to resume and pursue a fully-fledged academic career after a hiatus from academia lasting almost four years.

- Lidia De Wet, for language editing the document.

Joy, beautiful sparkle of the gods, Daughter of Elysium, We enter, fire-drunk, Heavenly one, your shrine. Your magics bind again, What custom has strictly parted, All men become brothers, Where your tender wing lingers.

(English translation of the first stanza of *An die Freude* by Friedrich Schiller (1785), unofficial lyrics to the anthem of the European Union, set to the Ode to Joy from Beethoven's 9th Symphony.)

ABSTRACT

The linear factor model is a building block of the Arbitrage Pricing Theory (APT). Macroeconomic factors may be used in linear factor models to proxy for the pervasive influences in returns. However, as the true return generating process is unobservable, macroeconomic data is either inaccurate or unavailable and because of the principle of parsimony, the linear factor model is likely to suffer from factor omission and consequent underspecification. Underspecification may adversely affect the interpretation of results, introduce coefficient bias, result in an upward bias in the residual variance and adversely affect predictive ability. The diagonality assumption that underlies the APT linear factor model will also be violated. Consequently, underspecification may pose a challenge to the general validity and interpretation of the linear factor model and the APT model. A widely applied solution to omitted factor bias in APT literature is the Burmeister and Wall (1986) residual market factor, hypothesised to fulfil the role of a wide-ranging proxy for omitted factors. This factor is derived from a broad market aggregate by excluding the influence of other factors that feature in a given linear factor model.

This study sets out to determine whether the use of a conventional residual market factor derived from a domestic market aggregate adequately resolves underspecification. This study also considers the impact of underspecification on the linear factor model. The role of a second residual market factor derived from a widely used global market index, the MSCI World Market Index, in resolving factor omission is also considered. A second residual market factor that is orthogonal by contribution to the factor set in the linear factor model should be irrelevant if a conventional residual market factor is an adequate proxy for omitted factors. Consequently, the second residual market factor in this study also fulfils the function of a test of the adequacy of the conventional residual market factor.

The approach in this study is comparative; three reduced form models are juxtaposed against a benchmark model and each other. The benchmark model incorporates a macroeconomic factor set, two residual market factors and a factor analytic augmentation as proxies for any remaining unobserved and omitted factors. Each specification is estimated using maximum likelihood (ML) estimation. Conditional variance is modelled as an ARCH(p) or GARCH(p,q) process to permit the structure of conditional variance to enter coefficient estimates and to provide insight into the conditional variance structure of the residuals. It is hypothesised that if factor omission has no impact on representations of the linear factor model and if the residual market factor is an effective and adequate proxy for omitted factors, then a model that comprises macroeconomic factors and a residual market factor should be comparable to the benchmark model in terms of results, general inferences and other aspects.

This study finds that a linear factor model incorporating only macroeconomic factors performs poorly. The significance of factors is understated and the model is misidentified. Standard errors and residual variance are inflated, coefficients are biased and predictive and explanatory performance is poor. Significant deviations from the true return generating process are observed and the diagonality assumption is violated. The incorporation of a single residual market factor improves such a specification although there is still evidence of significant omitted factor bias. Violations of the diagonality assumption continue to persist but are not as widespread as for the specification that solely employs macroeconomic factors. The inclusion of a second residual market factor does not significantly alleviate the symptoms of underspecification and this factor is significant in a number of instances suggesting that the residual market factor does not capture all omitted influences by itself.

Researchers of the APT and practitioners are encouraged to take note of these findings to avoid misinterpreting the results of macroeconomic linear factor models. The linear factor model is a complex construct and the application of a widely used approach in APT literature to resolve factor omission may not be adequate. This can adversely impact studies focusing on the linear factor model and equilibrium pricing within the APT and studies that apply macroeconomic linear factor models motivated by the APT.

Keywords: Arbitrage Pricing Theory, macroeconomic factors, linear factor model, return generating process, factor omission, underspecification, residual market factor, proxy

TABLE OF CONTENTS

LIST OF TABLES

TABLE OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Full Definition
APT	Arbitrage Pricing Theory
ASE	American Stock Exchange
AIC	Akaike Information Criterion
ASE	American Stock Exchange
ADF test	Augmented Dickey-Fuller test
ARCH	Autoregressive Conditional Heteroscedasticity/Heteroscedastic
BIC	Bayesian Information Criterion
BLU estimators (BLUE)	Best linear unbiased estimators
BP_t	Building plans passed (innovations)
BUS_t	Domestic business activity (innovations)
CAPM	Capital Asset Pricing Model
CRSP	Centre for Research in Security Prices
DJIA	Dow Jones Industrial Average
EU	European Union
GARCH	Generalized Autoregressive Conditional Heteroscedasticity
GDP	Gross Domestic Product
GNP	Gross National Product
HAC standard errors	Heteroscedasticity and autocorrelation consistent standard errors
IAPM	International Asset Pricing Model
IAPT	International Asset Pricing Theory
$IM\varepsilon_t$	Residual market factor derived from returns on the MSCI World Market Index (the international/second residual market factor)
JB test	Jarque-Bera test
JSE	Johannesburg Stock Exchange
$LEAD_t$	Composite leading business cycle indicator (innovations)
LR test	Likelihood Ratio test
LS	Least squares
LSPD	London Share Price Database
LSE	London Stock Exchange
$M\varepsilon_t$	Residual market factor derived from returns on the JSE All Share Index
ML	Maximum likelihood
MAP	Minimum Average Partial
MET_t	World metal prices in US Dollar (innovations)
MSCI	Morgan Stanley Capital International
NASDAQ	National Association of Securities Dealers Automated Quotation
NYSE	New York Stock Exchange
NRI	Nomura Research Institute
NLSUR	Non-linear seemingly unrelated regression
NL3SLS	Non-Linear Three Stage Least Squares regression
$\bar{R}^2$	Adjusted coefficient of determination

RESET test	Ramsey (1969) regression specification error test
S&P 500 Index	Standard & Poor's 500 Index
$TLI\varepsilon_t$	Composite leading conditions indicator for South Africa's trading partners (innovations, orthogonalised)
TSE	Toronto Stock Exchange or Tokyo Stock Exchange, depending upon context
UK	United Kingdom
US	United States
$USD\varepsilon_t$	Rand-Dollar exchange rate (innovations, orthogonalised)

LIST OF DEFINITIONS

Term	Definition
APT framework	A theoretical framework that permits the multifactor modelling of return behaviour, consisting of the linear factor model representative of the return generating process and the cross-sectional APT relation that relates expected returns to estimated factor coefficients.
APT relation/model	Model that relates expected (equilibrium) returns to factor coefficients derived from the linear factor model. Unlike the linear factor model which explains the time series behaviour of returns, the APT model explains cross-sectional differences in expected returns.
ARCH effect	Varying amplitude of returns or residuals over time indicating unequal variance.
Benchmark model	The benchmark model in this study comprises macroeconomic factors, a residual market factor derived from returns on the JSE All Share Index, a second residual market factor derived from returns on the MSCI World Market Index and a factor analytic augmentation (equation (6.20) and equation (8.2)). The restricted and unrestricted models are compared to this specification.
Chen, Roll and Ross (1986) factors/the conventional factors	The macroeconomic factor set studied in the seminal work on the macroeconomic APT by Chen, Roll and Ross (1986). These factors are the growth rate in monthly industrial production, the change in expected inflation, unexpected inflation, the risk premium and the term structure of interest rates.
Coefficient of determination ($\bar{R}^2$)	Measure of the proportion of variance in the dependent factor explained by the independent factors in a specification, which may be viewed as the explanatory power of a model.
Conditional heteroscedasticity	Heteroscedasticity that is dependent upon model specification as conditional variance is dependent upon the variance of the residuals derived from a given specification.
Conditional variance structure	Structure of the residual variance, as described by an ARCH(*p*) or GARCH(*p,q*) process.
Conventional/single residual market factor	Residual market factor derived by regressing the returns on a domestic/national market aggregate onto the factors that feature in a linear factor model specification.

Factor	General term for a variable that may be either pre-specified such as a macroeconomic variable or a variable that is statistical in nature. In this study, the term *factor* is used to refer to variables in the general sense as per classical APT literature.
Factor analysis	Technique that extracts sources of common variability and correlation from a set of factors/series and summarises these using a reduced number of factor scores.
Factor analytic augmentation	Inclusion of factor scores in a linear factor model, which are assumed to represent omitted and unobserved factors. Factor scores are derived from the residuals of a reduced form version of a given model.
Firm-specific	Analogous to industry-specific factors, depending upon context. These are factors that impact a specific firm or industry and are uncorrelated with systematic factors/risk and are diversifiable (see idiosyncratic factors)
Global/international market index	Index representative of movements on major global stock markets.
Heteroscedasticity	Unequal variance across observations of a specific series (see ARCH effect).
Idiosyncratic factors	Factors that have an impact that is limited to a specific asset. Analogously, firm-specific or sector-specific factors, depending upon context.
Impure heteroscedasticity	Heteroscedasticity in the residuals of a model arising as a result of factor omission.
Impure serial correlation	Serial correlation in the residuals of a model arising as a result of factor omission.
Innovations/unexpected components	The components of a factor that are unpredictable and exceed the expectations of economic agents.
Linear factor model	Linear representation of the return (data) generating process relating realised returns to innovations in systematic/pervasive factors over time.
Macroeconomic APT	APT relation characterised by pre-specified macroeconomic factors.
Macroeconomic linear factor model	Model relating realized returns to changes in pre-specified macroeconomic factors.
Orthogonalisation/ residualisation	Process of extracting the influence of other factors from a given factor so that the factor in question is uncorrelated with the orthogonalising factor set.
Pervasive/systematic/common factors	Factors that have an economy/market-wide impact. Used interchangeably.
Predictive ability	Ability of a model to accurately replicate actual return observations.

Priced factor	Factor for which the risk premium associated with a sensitivity to a factor in the linear factor model is statistically significant in the cross-sectional APT relation.
Pseudofactors	Factors for which the impact on returns is limited to a subperiod or subsample of the full sample.
Residual correlation	Pairwise correlation (alternatively cross-correlation) between the residual series derived from a linear factor model; referred to as residual correlation in the text (contextualised).
Residual serial correlation	Existence of a relationship between the residuals of a single series over time.
Residuals	Portion of a dependent factor that is not explained by a model. The residuals are assumed to reflect the impact of factors that have been omitted from a specification.
Restricted model	The restricted model in this study uses macroeconomic factors to characterise the linear factor model (equation (6.21) and equation (9.1)).
Return generating process	Unobserved process that drives movements in financial series.
Underspecification	Deliberate or involuntary omission of relevant factors from the linear factor model.
Unrestricted market model	The unrestricted market model in this study uses macroeconomic factors and a residual market factor derived from returns on the JSE All Share Index to characterise the linear factor model (equation (6.22) and equation (10.1)).
Unrestricted model	The unrestricted model in this study uses macroeconomic factors, a residual market factor derived from returns on the JSE All Share Index and a second residual market factor derived from returns on the MSCI World Market Index to characterise the linear factor model (equation (6.19) and equation (10.2)).
Unrestricted models/specifications	Collective term used to refer to both the unrestricted market model and the unrestricted model (set out above).

CHAPTER 1

INTRODUCTION

1.1. BACKGROUND

The Arbitrage Pricing Theory (APT) introduced by Ross (1976) can by summarised by two complementary specifications. The first is the linear factor model, which postulates that stock returns are generated by a multifactor model that is a representation of the return generating process. The second, often referred to in the literature as the APT relation, relates factor coefficients (the betas) derived from the linear factor model to expected returns to obtain risk premia associated with factors in the linear factor model. The nature of the APT emphasises the importance of the linear factor model which is a building block of the APT as a conceptual framework (Elton & Gruber, 1997: 1750; Drakos, 2002:74).

Early studies[1] of the APT rely upon factor analytic approaches to identify the number of factors in the return generating process and to derive factor betas for use in tests of the APT. However, factors derived using factor analysis are statistical in nature and uninterpretable, posing a major limitation for researchers and practitioners of the APT. Chen, Roll and Ross (1986) overcome this limitation by employing a set of macroeconomic factors to proxy for the unspecified pervasive influences represented by factors derived using factor analysis. Motivated by this early work, the macroeconomic APT has been applied to numerous markets and in different contexts (Cauchie, Hoesli & Isakov, 2004). Some of these studies investigate pricing by applying the APT relation. Other studies are of a non-equilibrium nature and relax some of the assumptions of the framework with the intention of estimating and studying an empirical linear factor model. Regardless of the manner of application, an accurate and correctly specified linear factor model is essential.

Van Rensburg (2000: 36) and Middleton and Satchell (2001: 506) argue that models that employ macroeconomic factors to describe the linear factor model are likely to be underspecified. Underspecification has consequences that extend beyond the immediate violation of certain assumptions underlying the APT framework, and specifically the linear factor model, and impact inference making and model parameters. An approach to resolving underspecification that finds favour in APT literature is the use of the Burmeister and Wall

[1] See Roll and Ross (1980), Chen (1983), Kryzanowski and To (1983), Hughes (1984), Beenstock and Chan (1986) and Elton and Gruber (1988) amongst others.

(1986) residual market factor derived from a market index. According to Berry, Burmeister and McElroy (1988: 31) and Elton, Gruber, Brown and Goetzmann (2014: 375), the residual market factor is a proxy for unobserved influences and resolves concerns over omitted factors. Burmeister and Wall (1986:10) use the residual market factor to capture “other market risk” not included in specifications of the linear factor model and Berry *et al*. (1988: 31) state that the concern over omitted factors is substantially resolved by including a residual market factor. Van Rensburg (1996:107) argues that the residual market factor “aids in alleviating fears of specification errors due to omitted variables.”

Given that the literature suggests that macroeconomic factors perform poorly in explaining returns (Connor, 1995; Van Rensburg, 2000), a residual market factor or even two residual market factors have been widely used to proxy for omitted factors in the literature. Studies which employ a residual market factor (or factors) are those of McElroy and Burmeister (1988), Koutoulas and Kryzanowski (1994), Clare and Priestley (1998), Brown, Hiraki, Arakawa and Ohno (2009) and Czaja, Scholz and Wilkens (2010). On the South African market, much work using a single residual market factor derived from returns on the FTSE/JSE All Share Index has been undertaken by Van Rensburg (1996; 1997). As an extension, Van Rensburg (2000; 2002) considers the use of two residual market factors derived from industrial indices as opposed to a residual market factor derived from a single broad market aggregate.

1.2. OBJECTIVES OF THE STUDY AND RESEARCH QUESTIONS

Underspecification of the macroeconomic linear factor model and by implication the macroeconomic APT relation is likely in general applications and APT pricing studies. There are numerous reasons for underspecification, namely the inability of macroeconomic factors to proxy for the true underlying pervasive influences, the unobservability and complexity of the return generating process, the unavailability of data, data inaccuracies, changes in the structure of the linear factor model and the principle of parsimony (Middleton & Satchell, 2001: 506; Gujarati, 2004: 45-46; Brauer & Gómez-Sorzano, 2004: 39; Spyridis, Sevic & Theriou, 2012: 55).

This study investigates whether the conventional residual market factor, derived from a broad domestic market aggregate, sufficiently addresses the consequences of underspecification and is therefore an adequate catch-all proxy for omitted factors. It also considers the impact of underspecification on the linear factor model when pre-specified macroeconomic factors are used as sole proxies for the pervasive influences in stock returns. A single residual market factor is first considered to determine whether the symptoms of underspecification are adequately mitigated in such a specification. In doing so, the econometric and theoretical impact of underspecification on the linear factor model is considered and quantified and indirectly related to the broader APT framework.

The study then introduces an extension by incorporating a second residual market factor in the form of an international residual market factor derived from the Morgan Stanley Capital International (MSCI) World Market Index, a commonly used international market index (Clare & Priestley, 1998; Brown *et al.,* 2009). This extension serves as a further test of the adequacy of the residual market factor as a proxy for omitted factors. If the conventional residual market is an adequate proxy for omitted factors, a second residual market factor, which is orthogonal to the macroeconomic factor set and the first residual market factor, should be redundant. By introducing a second residual market factor, this study considers whether two residual market factors may be more appropriate for capturing the impact of omitted factors. The literature recognises that global macroeconomic factors play an important role in determining stock returns and that returns depend upon the level of market integration. The use of this specific index, the MSCI World Market Index, is in keeping with convention in the literature, as is the use of a broad domestic market aggregate to derive the conventional residual market factor (Ferson & Harvey, 1994; Harvey, 1995; Bilson, Brailsford & Hooper, 2001; Szczygielski & Chipeta, 2015). That international factors play a role in driving returns is especially true for emerging markets that have experienced reductions in barriers to capital mobility and ongoing globalisation since the 1990s (Clare & Priestley, 1998: 104).

Following from the above discussion and in summary, the research questions explored in this study are:

1) The ability of the pre-specified macroeconomic factors to act as proxies for pervasive influences in stock returns;
2) The impact of factor omission, namely underspecification, on the linear factor model within the context of the APT framework;
3) The efficacy of using a conventional residual market factor to resolve underspecification; and

4) The efficacy and relevance of a second residual market factor in resolving any extant underspecification.

1.3. METHODOLOGY

This study follows a comparative research design in that it seeks to identify and quantify the consequences of factor omission on the linear factor model by comparing different specifications of a linear factor model to a well-specified benchmark model.

The factor structure of the South African stock market is explored to identify macroeconomic factors that are proxies for pervasive influences in returns. In identifying macroeconomic factors and confirming that these factors are proxies for pervasive influences in stock returns, an approach motivated by the work of Chen and Jordan (1993), Choi and Rajan (1997), Panetta (2002) and Spyridis *et al.* (2012) is followed. Statistical factors derived from returns on South African industrial sectors for the period January 2001 to December 2016 are correlated with a limited set of pre-specified factors selected from a broader set of candidate macroeconomic factors. Macroeconomic factors that are found to proxy for the pervasive influences in returns are incorporated into four specifications. The first is a benchmark model. This model incorporates the pre-specified macroeconomic factor set, two residual market factors and a factor analytic augmentation that represents unobserved and omitted factors (Van Rensburg, 1997: 63). The benchmark specification is hypothesised to represent an adequately specified linear factor model, which is free of underspecification. The next model is a restricted version of the benchmark specification, which incorporates only the macroeconomic factors identified as proxies for pervasive influences. The unrestricted market model incorporates the macroeconomic factors and a residual market factor derived from returns on the JSE All Share Index. The unrestricted specification incorporates these factors and a second residual market factor derived from returns on the MSCI World Market Index. Collectively, in this study, the latter two models, the unrestricted market model and the unrestricted model, are referred to as the unrestricted models/specifications. To estimate these specifications, maximum likelihood (ML) estimation is applied and conditional variance is modelled using Engle's (1982) and Bollerslev's (1986) Autoregressive and Generalized Autoregressive Conditional Heteroscedastic (ARCH(p) and GARCH(p,q)) models. The use of this econometric framework not only produces efficient coefficient estimates, but also permits the impact of underspecification to be reflected in the conditional variance structure and coefficient estimates by permitting information in the residual variance to enter coefficient estimates (Hamilton, 2010; Brzeszczyński, Gajdka & Schabek, 2011: 33).

To investigate the impact of underspecification on the linear factor model, the restricted model that incorporates macroeconomic factors, is juxtaposed against the benchmark specification. To investigate the ability of the residual market factors to resolve underspecification, the unrestricted models are juxtaposed against the benchmark and the restricted model. Comparisons are made across numerous aspects of the models to identify and describe the impact of factor omission. These are factor significance, coefficient bias, the accuracy of inferences and interpretation, explanatory power, the ability of the models to approximate the true return generating process, model diagnostics, residual variance bias, the structure of the conditional variance and predictive ability (discussed in Chapter 5). Importantly, the structures of the respective (pairwise) residual correlation matrices derived from each specification are considered to determine whether these matrices conform to the assumption that residuals are uncorrelated across series (the diagonality assumption), which underpins the linear factor model. It is hypothesised that if the conventional residual market factor adequately resolves underspecification by accounting for omitted factors in addition to a macroeconomic factor set, then the abovementioned aspects of a model that incorporates a residual market factor should be comparable to those of the benchmark model and no other orthogonal factors should be relevant. Also, a comparison of the unrestricted models to the restricted model indicates whether the use of a residual market factor goes some way to resolving factor omission bias.

1.4. CONTRIBUTION OF THE STUDY

The essence of this research is the study of the impact of factor omission on the linear factor model and the ability of the residual market factor to adequately resolve underspecification. Such an investigation is also an investigation of the validity of the diagonality assumption that underlies the linear factor model. This assumption should hold if macroeconomic factors and a single residual market factor or even two residual market factors are able to adequately account for common movement in returns (Elton & Gruber, 1988: 31; Elton *et al.,* 2014: 157). Van Rensburg (1997: 59; 2002) argues that the assumption of uncorrelated residuals, in other words, of adequate specification, is often neglected yet is of serious concern and consequence. In the context of the APT, Elton, Gruber and Blake (1995: 1239) state that the APT will fail if the return generating process is misspecified.

This study deals with the lacuna set out by Van Rensburg (1997; 2002) by comprehensively investigating the theoretical and econometric impact on the APT linear factor model of the violation of the diagonality assumption attributable to underspecification. This is investigated within the broader APT framework and the study is faithful to the techniques employed in APT literature in identifying, constructing and estimating the linear factor model. This study highlights the inadequacy of macroeconomic factors and the challenge associated with using macroeconomic factors to describe the linear factor model. This study also sets out to determine whether a widely used approach, that of employing a residual market factor (or multiple residual market factors), sufficiently resolves concerns over omitted factors.

This study's contribution is also methodological. Conventional statistical theory postulates that the least squares coefficients of a model are unbiased and consistent but not efficient in the presence of heteroscedasticity (Armitage & Brzeszczyński, 2011: 1526). However, Hamilton (2010: 20) argues that if heteroscedasticity is not corrected for, least squares and ARCH model parameters (in the conditional mean) will differ significantly. Bera, Bubnys and Park (1988: 209, 211-212) state that by the nature of the ARCH process, the impact of omitted factors will be reflected in the model coefficients of the conditional mean equation and that the magnitude of the impact will be dependent upon the level of conditional heteroscedasticity, which, in turn, is dependent upon factor omission (Webster, 2013: 230). By applying ARCH(p) and GARCH(p,q) models to model the conditional variance underlying the linear factor models estimated, this study investigates the impact of underspecification on the structure of the conditional variance and also contributes to literature that acknowledges and quantifies the impact of heteroscedasticity and the conditional variance structure on coefficient estimates.

The findings of this study are that a factor analytic augmentation should be considered even when a residual market factor or two residual market factors are used to proxy for omitted factors. Specifically, versions of the linear factor model that incorporate macroeconomic factors, and subsequently, a single residual market factor, and then finally, a second residual market factor, underperform a model that comprises these factors and a factor analytic augmentation. The factor analytic augmentation represents unspecified and unobserved factors relegated to the residuals of the linear factor model, which in APT literature, are widely assumed to be reflected by the residual market factor in the first instance (Elton *et al.,* 2014: 375; Berry *et al.,* 1988: 31). Specifications that rely on macroeconomic factors and also incorporate a residual market factor to resolve factor omission are prone to erroneous interpretations, suffer from coefficient bias, inflated variance, larger prediction errors and poor predictive performance. A failure to account for relevant omitted factors, some of which appear to be unspecified and unobserved as suggested by the presence of statistical factors derived from the residuals of the unrestricted models, results in a violation of the diagonality assumption. Moreover, the residual correlation matrix of the restricted model exhibits strong pairwise interdependence, which is reduced by the inclusion of the residual market factors, suggesting that the levels of pairwise residual correlation are indicative of factor omission. Finally, it appears that residual correlation diagonality is a somewhat restrictive assumption, which does not hold in practice and is at best a desirable approximation of reality (Elton *et al.,* 2014: 157). The diagonality assumption will be violated for even a well-specified benchmark linear factor model. However, the pairwise residual correlation for such a model is generally low in magnitude and correlations do not exhibit systematic patterns.

The findings are of interest to researchers and practitioners of the APT and to researchers who employ more generic and less theory-driven approaches to modelling financial markets on the basis of macroeconomic factors and the general macroeconomic state. An additional approach, as far as possible, to resolving underspecification, is suggested, namely that of using a factor analytic augmentation in linear factor models.

1.5. OUTLINE OF THE STUDY

Chapter 2 presents a developmental overview of the APT framework by setting out its underlying assumptions, reviewing its development, limitations and applications. Chapter 3 introduces the residual market factor, outlines its theoretical basis and reviews its application in the literature. Chapter 4 discusses the role of international influences in financial markets and emphasises their importance, contextualises their role in APT literature and in doing so, motivates for the candidacy of an international market index for a second residual market factor. Chapter 5 outlines and demonstrates the consequences of underspecification and relates these to APT theory and application. The data and the methodology which are employed in investigating underspecification of the linear factor model and the efficacy of the residual market factor in resolving underspecification are outlined in Chapter 6. Chapter 7 investigates the factor structure underlying the South African stock market and identifies macroeconomic factors that proxy for pervasive influences in stock returns.

Chapter 8 constructs the benchmark specification that incorporates macroeconomic factors, the two residual market factors and a factor analytic augmentation. The results of the model are interpreted to show that the estimated model meets *a priori* expectations relating to the hypothesised impact of macroeconomic factors in the model. The resultant residual correlation matrix is factor analysed and it is shown that this specification is adequately specified. Chapter 9 reports the results of the estimation of the restricted model that incorporates the macroeconomic factors identified in Chapter 7. Model parameters and various aspects of the model are compared to those of the benchmark specification to determine how these are impacted by the omission of the residual market factors and the factor analytic augmentation. Chapter 10 reports on the two unrestricted models, namely the unrestricted market model, which incorporates the residual market factor in addition to the macroeconomic factors, and the unrestricted model, which incorporates a second residual market factor in addition to the factors in the unrestricted market model. The parameters of these models and other aspects are compared to those of the restricted and the benchmark model to determine whether the inclusion of the residual market factors produces an improvement over the restricted model and whether these models approximate the benchmark model. Chapter 11 concludes by summarising the findings and implications of the study, suggests reasons for the observed results and sets out avenues for further research.

1.6. DELIMITATIONS

The purpose of the study is to investigate the impact of underspecification on the linear factor model and the adequacy of a residual market factor to sufficiently resolve factor omission. While it is acknowledged that the linear factor model is a building block of the APT, the focus is on the macroeconomic linear factor model as the basis of a descriptor of the return generating process. The focus is not on the impact of underspecification on the APT relation and the associated pricing implications, although these aspects are acknowledged and accordingly discussed. Also, this study does not seek to investigate the reasons as to why macroeconomic factors and residual market factors fail to adequately describe the return generating process. Reasons are suggested in the conclusion but are not explored further and this is treated as an avenue for further research.

This study does not seek to investigate the appropriateness of a specific market index as a domestic or international market proxy[2] from which the residual market factors are derived. In line with application in the literature, the domestic aggregate, the JSE All Share Index, is used to derive the domestic residual market factor and the MSCI World Capital Market Index is used to derive the international residual market factor (McElroy & Burmeister, 1988; Clare & Priestley, 1998; Bilson *et al.,* 2001). Although more appropriate indices may be constructed, the intention of this study is to consider whether commonly used indices, in the generic sense and of which the application is practical, are sufficient to derive residual market factors that are adequate proxies for omitted factors. This study acknowledges that the diagonality assumption is an approximate assumption underlying the linear factor model and that it may not hold in practice. Consequently, the diagonality assumption is treated as a theoretical construct that may be unachievable in practice. Therefore, this study does not attempt to construct a model that will produce a residual correlation matrix that conforms to this assumption. Rather, this study seeks to construct a benchmark model that will produce an acceptable practical approximation of this assumption.

The modelling of conditional variance structures is restricted to two specifications, the ARCH*(p)* and GARCH*(p,q)* specifications. Given that conditional variance is characterised by a number of stylised facts such as asymmetry and long-memory, there may be other ARCH/GARCH-type specifications, such as Nelson's (1991) Exponential GARCH(*p,q,n*) model, which are more appropriate (Xiao & Aydemir, 2007: 3). Because the impact of underspecification on the structure of the conditional variance is only one of the aspects considered, the appropriateness of other such specifications and their relation to underspecification is an area for further research. Finally and most importantly, this study should not be seen as a challenge to the validity of the linear factor model that underpins APT. Rather, it is an investigation of whether macroeconomic factors can characterise the linear factor model and a test of whether the conventional approach of using a residual market factor adequately resolves underspecification.

[2] See Brown and Brown (1987) for a discussion of the composition of a market proxy and the consequences of differing the composition of a hypothetical market proxy.

CHAPTER 2

A DEVELOPMENTAL OVERVIEW OF THE ARBITRAGE PRICING THEORY

2.1. INTRODUCTION

The purpose of this chapter is to provide an overview of the APT framework and the role of the linear factor model within the APT framework. The development of the APT is set out and its international extension is introduced. The overview culminates in a discussion of the linear factor model and the corresponding macroeconomic APT and its application in asset pricing and linear factor model literature.

The essence of the APT is encapsulated by two separate yet complementary equations; the linear factor model and the APT model. The linear factor model is the basis for time series multifactor specifications of the return generating process and the APT model relates returns in equilibrium (expected returns) to exposures to factors in the linear factor model. Although the two models can be seen as serving two different purposes, both models are intimately linked. The linear factor model is a building block of the APT model and is used to derive inputs for the APT model whereby estimates of sensitives to factors in the linear factor model are used as "data" in the APT model (McElroy & Burmeister, 1988: 31; Elton & Gruber, 1997: 1750; Drakos, 2002: 74). For this reason, the linear factor model is an important construct in the APT framework and is the focus of study. Nevertheless, an understanding of the APT is crucial as the APT is underpinned by the linear factor model and requires an appropriate specification of the linear factor model.

This chapter proceeds by first providing an overview of the APT framework and the role of the linear factor model within the framework in Section 2.2. The development of the APT framework is outlined in Section 2.3. by reviewing early studies, introducing an extension in the form of the international APT and by acknowledging the early limitations of the APT noted in the literature. One of these limitations, namely the use of uninterpretable statistical factors, hinders interpretability and spurs the development of the macroeconomic APT as a response to criticisms. The development and application of the macroeconomic APT are discussed in Section 2.3.4. This is followed by a discussion and overview of the applications of the macroeconomic APT in Section 2.4., with the emphasis being on the multifactor character of the macroeconomic APT. The conclusion in Section 2.5. summarises the chapter and draws inferences from the literature reviewed.

2.2. AN OUTLINE OF THE APT

The genesis of the APT begins with Ross (1976: 341), who traces its origin as an alternative to the Capital Asset Pricing Model (CAPM) introduced by Treynor (1961), Sharpe (1964) and Lintner (1965). Roll and Ross (1980: 1073) state that although the CAPM has dominated empirical work in the past and is the basis of modern portfolio theory, research has brought into question the ability of the CAPM to explain the behaviour of returns. Examples of such studies are those of Basu (1977), Banz (1981) and Fama and French (1992), who suggest that expected returns are related to firm-specific factors such as size, the earnings-to-price and book-to-market ratios as opposed to a sole hypothesised determinant of cross-sectional returns, the market beta, β_M, representative of market risk (Keim, 1986 : 20; Burmeister, Roll & Ross, 1994: 3). Research has also brought into question the foundation of the CAPM, the diagonal model introduced by Sharpe (1963), a single-factor model of the return generating process, which Roll and Ross (1980: 1074) view as the "intuitive grey eminence behind the CAPM." Examples of such studies are those of King (1966), who finds support for two common sources of variation in returns on stocks listed on the New York Stock Exchange (NYSE), namely market and industry components, and that of Meyers (1973), who challenges King's (1966) results but at the same time recognises that there may be numerous unexplained components in returns that could challenge the validity of a single-factor model. These challenges to the validity of the CAPM and its underlying foundation, have led to the APT framework being proposed and developed as a testable successor to the CAPM and the underlying single-factor model (Chen, Hsieh, Vines & Chiou, 1998: 279-280).

The point of departure for expounding the APT is the linear factor model, which, in contrast to the CAPM and the single-factor diagonal model, permits for more than a single return generating factor to feature in the return generating process (Sadorsky, 2008: 3855; Elton & Gruber, 2018: 98):

$$R_{it} = E(R_i) + \sum_{k=1}^{K} b_{ik} f_{kt} + \varepsilon_{it} \tag{2.1}$$

where R_{it} is the return on asset *i* at time *t*, $E(R_i)$ is the expected return on asset *i*, f_{kt} is a realisation of the *k*th factor at time *t*, which is assumed to impact realised returns, and b_{ik} is the sensitivity of asset *i* to factor *k*. As investors can diversify by holding portfolios of assets, only systematic (pervasive/common) factors drive returns on assets and all idiosyncratic effects cancel out and are relegated to the residuals, ε_{it} (Burmeister *et al.,* 1994: 3). Aside from factors being of a systematic nature, the APT also requires that each factor is unpredictable at the beginning of each period, that is $E(f_{kt}) = 0$, and is associated with a non-zero risk premium in the APT relation (equation (2.4), Berry *et al.,* 1988: 30). According to the linear factor model, realised returns are a function of an asset's expected return, realisations in relevant systematic factors and returns attributable to asset-specific events (Roll & Ross, 1995: 197).

The residual error terms, ε_{it} in equation (2.1), warrant further attention and are important to this study. Two assumptions are made relating to the nature of the residuals within the APT framework, summarised as follows:

$$\text{cov}(\varepsilon_{it}, \varepsilon_{jt}) = 0 \tag{2.2}$$

$$\text{cov}(\varepsilon_{it}, f_{kt}) = 0 \tag{2.3}$$

Equation (2.2) represents the assumption of uncorrelated residuals for assets *i* and *j,* and can also be denoted by $E(\varepsilon_{it}, \varepsilon_{jt}) = 0$. This assumption implies that the reason that assets move together is because of co-movement associated with a set of common factors that should be reflected in the linear factor model. If this assumption does not hold, then this implies that there are other factors in the linear factor model than the *k* hypothesised factors in the factor set, $\sum_{k=1}^{K} b_{ik} f_{kt}$, in equation (2.1), which explain any remaining residual co-movement. The violation of this assumption represents a specification error attributable to the omission of relevant factors – underspecification (Van Rensburg, 2000: 36). The second assumption set out in equation (2.3) is that the residuals of the linear factor model are uncorrelated with factor realisations, as denoted by $E(\varepsilon_{it}, f_{kt}) = 0$ (Burmeister *et al.,* 1994: 4, Elton & Gruber, 2018: 98). As with equation (2.2), the assumption set out by equation (2.3) will also be violated if the factor set in equation (2.1) fails to reflect all relevant systematic factors. The influence of relevant but omitted factors will be relegated to the residuals and the residuals will be correlated with factors in the factor set and with the omitted factor(s)

(Studenmund, 2014: 179). This study is concerned with $E(\varepsilon_{it},\varepsilon_{jt})=0$; it follows that if this assumption does not hold, $E(\varepsilon_{it},f_{kt})=0$ may not hold.

The APT relation completes the description of the APT framework and relates expected returns to the factor coefficients estimated within the linear factor model (Burmeister *et al.,* 1994: 5):

$$E(R_i)=\lambda_0+\sum_{k=1}^{K}\lambda_k b_{ik} \tag{2.4}$$

where $E(R_i)$ is the expected return on asset *i*, λ_0 is the return on a riskless asset (an asset that is not exposed to systematic risk if such an asset exists), λ_k is the risk premium associated with an exposure, as represented by b_{ik}, to factor f_{kt} estimated in equation (2.1), the linear factor model. A factor is said to be priced if a risk premium associated with a certain factor is statistically significant. The risk premia are indicative of compensation for exposure to systematic factors in the linear factor model (Elton & Gruber, 1988: 42) . The APT model, as exemplified by equation (2.4), arrives at the following outcome; the cross-section of expected returns is determined by the expected return on a riskless asset and an asset's sensitivities to common factors that feature in the linear factor model and the associated risk premia (Reinganum, 1981: 314).

The foundational role of the linear factor model in the APT is perhaps best demonstrated by Fama and MacBeth (1973), whose two-step approach is deemed to be the standard (and also an early) procedure applied in empirical tests of the CAPM and (later) the APT. This technique is used to establish whether the factors that feature in the linear factor model are true APT factors (Chimanga & Kotze, 2009: 83). In the first step, sensitivities(the coefficients), the b_{ik}s in equation (2.1), to systematic factors are estimated in time series regressions (the linear factor model). In the second step, risk premia, λ_k, are estimated using cross-sectional regressions of returns for each time period on the b_{ik}s. Averages of λ_k are then calculated over the whole sample period to obtain the risk premia (French, 2017: 13). Therefore, in contrast to the linear factor model which is a time series model of realised returns, the APT model is a cross-sectional model that establishes the equilibrium relationship between returns and the exposure to risk (Fama & MacBeth, 1973; Amenc & Le Sourd, 2003: 150, 192; Yao, Mei & Clutter, 2014: 945). Elton *et al.* (1995: 1239) aptly summarise the importance of the linear factor model in the APT by stating that the APT model will fail if the linear factor model is misspecified. This emphasises the importance of the linear factor model and provides impetus for the study of the consequences of underspecification for the linear factor model within the context of the APT and the ability of the residual market factor to resolve factor omission.

Finally, the APT is underpinned by a number of assumptions, which are summarised as follows (Basu & Chawala, 2012: 422):

1) Pure arbitrage profits are impossible as a result of competitive and frictionless markets and therefore a positive expected rate of return can only be earned by taking on risk and making a net investment of funds.
2) Investors are risk averse wealth maximisers.
3) Investors have homogeneous beliefs regarding the structure of the return generating process.

The APT is more general and less restrictive than the CAPM framework; the APT does not specify the factors that feature in the return generating process and which are priced or how many such factors exist. Furthermore, the assumptions that are required for the development of the CAPM, namely that 1) investors possess a quadratic utility function, 2) returns are normally distributed, and 3) a mean-variance efficient portfolio, are not required under the APT framework (Reilly & Brown, 2012: 148, 242).

The APT framework can be elegantly summarised by substituting equation (2.4) into equation (2.1). This yields what Berry *et al.* (1988: 31) refer to as the "full APT":

$$R_{it} = \lambda_0 + \sum_{k=1}^{K} \lambda_k b_{ik} + \sum_{k=1}^{K} b_{ik} f_{kt} + \varepsilon_{it} \tag{2.5}$$

It is perhaps this elegance that supports Roll and Ross' (1980: 1074) assertion that the APT framework's "modest assumptions and its pleasing implications surely render the APT worthy of being the object of empirical testing."

2.3. THE DEVELOPMENT OF THE APT FRAMEWORK

2.3.1. Early Studies

Early studies of the APT focus on testing the central tenets of the framework. These are those of a multifactor return generating process, the pricing of multiple factors and the relevance of systematic factors as opposed to firm-specific factors (henceforth broadly referred to as idiosyncratic factors). Also considered is the superiority of APT in explaining returns over time and in the cross-section relative to the CAPM.

Roll and Ross (1980: 1088, 1092) are amongst the first to empirically test the central propositions of the APT using portfolios of stocks listed on the New York Stock Exchange (NYSE) and the American Stock Exchange (ASE). Using factor analysis, Roll and Ross (1980) find that a five factor structure is sufficient to describe the return generating process and report that there are at least three but no more than four priced factors in expected returns. To test the validity of the APT, the authors investigate whether own standard deviation (a hypothesised asset-specific factor) is priced in each portfolio. Results indicate that the impact of own standard deviation on expected returns is limited to just three out of 42 portfolios. This supports the proposition of the APT that expected returns are explained by systematic factors and not asset-specific factors. Kryzanowski and To (1983) investigate the number of factors in the return generating process using return data for stocks on the NYSE and the Toronto Stock Exchange (TSE). At least 10 factors feature in the linear factor model describing returns on the US stock market and between 18 to 20 factors feature in the linear factor model underlying the Canadian stock market. However, factors beyond the fifth factor account for a low percentage of common variation suggesting that a five factor structure is sufficient from an economic perspective (Kryzanowski & To, 1983: 44, 48). These findings are similar to those of Roll and Ross (1980) and provide empirical support for a linear factor model that is characterised by multiple factors, although Roll and Ross (1980) also provide support for the multifactor APT relation.

Chen (1983: 1400) imposes a five factor structure onto returns for stocks in the Centre for Research in Security Prices (CRSP) database. Results indicate that between two and four factors are priced for most subperiods, with the exception of a single subperiod for which a single factor is priced.[3] A test of joint significance of the estimated risk premia in the five factor cross-sectional APT relation indicates that the risk premia are jointly statistically significant. From these results, it can be inferred that the linear factor model is characterised by multiple factors which are reflected in the APT relation. To compare the cross-sectional

[3] The subperiods considered are 1963 to 1966, 1967 to 1970, 1971 to 1974 and 1975 to 1978. Only one risk premium is significant for the 1971 to 1974 period. More than one risk premium is significant for all other subperiods

explanatory power of the APT and the CAPM, the Davidson and Mackinnon (1981)[4] test is applied and the results favour the APT. This indicates that a multifactor pricing relation outperforms a single-factor relation. Chen (1983) does not find evidence of an own variance or a firm size effect; there are no statistically significant differences between average returns for high-low variance portfolios and large-small firm portfolios. This suggests that systematic risk is fully accounted for by factor loadings, as proposed by the APT, and therefore diversifiable factors do not explain returns. Bower, Bower and Logue (1984: 1046) extend the investigation of the explanatory power of the APT beyond that of the APT relation and consider the explanatory power of a multifactor linear factor model. Using data for industrial portfolios formed from stocks listed on the NYSE and the ASE, the authors find that a (statistical) four-factor model explains a greater proportion of return variation relative to a market model based upon a CRSP value-weighted index (respective coefficients of determination, $\bar{R}^2$, are 0.869 and 0.605). In the cross-section, the APT relation explains almost double the variation explained by the CAPM (respective $\bar{R}^2$ of 0.425 vs 0.274). The linear factor model again outperforms the single-factor market model in explaining time series variation in a hold-out sample consisting of utility stocks, as evident from the respective $\bar{R}^2$s (0.323 and 0.263). The authors also report that the APT outperforms the CAPM in explaining the cross-sectional variation in returns, as evident from lower unexplained variances derived for the APT. Finally, Theil's *U* statistics are applied to assess the ability of the APT and CAPM to forecast expected returns. The APT outperforms the CAPM in forecasting returns for the hold-out sample. Chen's (1983) findings confirm that multiple factors are priced in the APT relation and that the multifactor APT outperforms the single-factor CAPM in explaining the cross-section of returns. Bower *et al.* (1984) demonstrate that specifications motivated by the APT outperform specifications informed by the CAPM across time and cross-sectionally in numerous aspects.

Brown and Weinstein (1983: 722, 724) investigate the hypothesis that common factors are responsible for determining returns using data from the CRSP database. Factor analysis carried out on 42 groups of stocks that comprise the sample indicates a three-factor model structure. Groups are then paired and an *F*-test is used to determine whether the same set

---

[4] The specification for the Davidson and Mackinnon (1981) test is $r = \alpha\hat{r}_{APT} + (1-a)\hat{r}_{CAPM} + \varepsilon$. Returns predicted by the APT relation are denoted by $\hat{r}_{APT}$ and returns predicted by the CAPM relation are predicted by $\hat{r}_{CAPM}$. Actual returns are denoted by *r*. An α of one or close to one indicates that the APT is the more appropriate model.

of factors generates returns for all stocks in a given group pairing. The null hypothesis that the same three factors describe the return generating process is rejected only four times indicating that the factors that best describe return behaviour do not significantly differ across groups – they are common and pervasive in nature. Hughes (1984: 207-208) finds that at least 12 factors explain returns for two groups of Canadian stocks and that the first factor is the most important and accounts for almost a third of the variation in returns. Factors beyond the fifth factor for both groups exhibit a trivial amount of explanatory power indicating that a five-factor model is sufficient. Between three and four factors are priced, a finding similar to that of Roll and Ross (1980).[5] To test whether the factors extracted are common across groups, expected returns for a given group are regressed onto factors extracted from the alternative groups. Between two and five factors are statistically significant across groups indicating that there is an underlying consistency across factor sets extracted from these two groups and used in pricing across alternate groups. Brown and Weinstein's (1983) and Hughes (1984) studies provide support for yet another tenet of the APT; that it is common factors that explain returns and are relevant in pricing.

Beenstock and Chan (1986: 128, 135-136, 138) investigate the factor structure of UK stocks and report that, depending upon sample period and criterion used, the number of factors lies between 19 and 29. In-sample tests of the APT indicate that in most instances, between zero and two factors are priced.[6] In comparisons of the cross-sectional explanatory power of the APT and the CAPM, the APT outperforms the CAPM both in-sample and out-of-sample (average $\bar{R}^2$ s of 0.263 vs 0.009 in-sample and 0.183 vs 0.023 out-of-sample). As in Chen (1983), the Davidson and Mackinnon (1981) test is applied, testing the CAPM against the APT. The results overwhelmingly favour the APT, both in-sample, out-of-sample, across periods and across sub-samples. Beenstock and Chan (1986) also investigate the presence of a firm-size effect by regressing firm size onto factor coefficients to establish whether the explanatory power of the factor coefficients is associated with firm-size. Firm size is found to be insignificant over the two subperiods and the three samples considered. As with other studies discussed, these findings provide general support for the APT.

---

[5] Hughes (1984) undertakes testing on Canadian stocks divided into two groups, Group A and Group B.

[6] Beenstock and Chan's (1986) sample spans the period between December 1961 and December 1981. It is then subdivided into two subperiods, 1962 to 1971 and 1972 to 1981. Stocks investigated within each subperiod are then sub-divided into three samples. For the 1961 to 1971 period, the number of priced factors ranges between zero and two, depending upon the type of statistical test applied (one-tailed or two-tailed) and for the 1972 to 1981 period, the number of priced factors ranges between zero and two.

Elton and Gruber (1988: 32, 43) investigate the factor structure of the Japanese stock market as represented by four samples formed from stocks that are part of the Nomura Research Institute (NRI) 400 index. A four factor structure is deemed sufficient to characterise the return generating process although cross-sectional tests indicate that only one factor is priced in returns on 20 portfolios formed from index constitutents. This latter finding is in contrast to the findings of the above-cited studies that report that expected returns generally reflect multiple sources of risk. Nevertheless, these results still support a multifactor return generating process. Similarly to Hughes (1984) and Brown and Weinstein (1983), Elton and Gruber (1988) also investigate the commonality of the factors derived from four samples formed from stocks constituting the NRI 400. It is shown that each factor derived from a single sample of stocks is highly correlated with the corresponding factor derived from a second sample.[7] The average correlation of each of the four factor solutions with a value-weighted index of the first section of the Tokyo Stock Exchange and 20 size portfolios formed from NRI 400 stocks is investigated next. For the value-weighted index, all correlations are above 0.90 and for the size portfolios, all average correlations are around 0.76. The comparability of the average correlation levels suggests that each four factor set is capturing the same influences and provides support for the commonality of the extracted factors. Finally, the authors report that a four- factor model outperforms a single index model that regresses returns on 20 portfolios onto returns on the NRI 400 in terms of explanatory power (average $\bar{R}^2$ of 0.77 vs 0.55.)

Yli-Olli and Virtanen (1992: 515) investigate the APT using Finnish stock return data quoted on the Helsinki Stock Exchange (HSE). Scree tests and transformation analysis are applied, providing support for a three to four stable factor solution across the two samples and three subperiods considered.[8] In cross-sectional tests, between one and two factors are priced, depending upon the subperiod considered. To investigate the validity of the APT, the residuals of three and four-factor APT relations are regressed onto own-variance and firm size. For a three-factor model, a significant own-variance effect is observed only for a single subperiod and a weak size effect is also observed for a single subperiod. The own variance effect disappears when the residuals of a four factor cross-sectional regression are used in

[7] For example, for a four factor solution, Elton and Gruber (1988: 28) report that the correlation between the first factor derived from Sample 1 and from Sample 2 is 0.974. All diagonal correlation coefficients for factors one to three derived from Sample 1 and Sample 2 are around 0.900. All off-diagonal elements are below 0.1.
[8] These are 1970 to 1975, 1976 to 1980, 1981 to 1986.

cross-sectional tests, indicating that the inclusion of an additional factor eliminates any residual systematic risk that might be reflected in own-variance.[9]

The literature generally supports the basic and central tenets of the APT. There is support for the proposition of a linear factor model that reflects multiple factors, postulated by the APT. Factors that feature in the linear factor model are priced in the APT relation and therefore explain returns in equilibrium. Usually, more than one factor is priced in cross-sectional returns although this is not always the case. Multifactor specifications motivated by the APT are also shown to be more adept at explaining the cross-sectional and time series behaviour relative to single-factor specifications. The literature also confirms that the factors that characterise the linear factor model are pervasive (systematic/common) in nature. Finally, as proposed by the APT, idiosyncratic factors do not appear to be important in pricing although the findings of Yli-Olli and Virtanen (1992) indicate that idiosyncratic factors may (misleadingly) proxy for omitted systematic factors (also see Fama & French, 1993; Brennan, Chordia & Subrahmanyam, 1998).

2.3.2. The International APT

In parallel with the testability of the proposition of the APT, the general form of the APT lends itself to extensions.[10] One such significant early extension is the international APT. The groundwork for this extension is set out in two papers by Solnik (1974; 1983).

Prior to the introduction of the APT as a framework, Solnik (1974: 365, 368) argues that the most realistic description of stock prices is provided by a multifactor model that takes into account both international and domestic factors. Solnik (1974) goes onto derive the International Asset Pricing Model (IAPM), analogous to the CAPM, in which international systematic risk is encapsulated by the beta on a world equity index. Preliminary tests for 10 developed markets[11] show that movements in a domestic market index explain between 16% and 46% (average $\bar{R}^2$ of 0.332) of variation in country returns. However, global

[9] See Section 5.4.2., Sweeney and Warga (1986: 398-399), Lehmann (1990: 72), Fama and French (1993: 8) and Brennan, Chordia and Subrahmanyam (1998: 349). It follows that idiosyncratic factors may proxy for omitted systematic factors. It also follows that if systematic factors fully account for risk in returns, idiosyncratic factors will no longer proxy for omitted systematic factors.

[10] This discussion is part of a review of the early development of the APT. The role of international influences in returns and in the broader APT framework is discussed in greater detail in Chapter 4.

[11] Solnik's (1974) sample encompasses stocks for France, Italy, the United Kingdom, Germany, the Netherlands, Switzerland, Belgium, Sweden, the US and Japan.

influences also explain a significant amount of variation in domestic returns; movements in the world index explain between 9% and 21% of variation in returns (average $\bar{R}^2$ of 0.153). Pricing tests support the validity of both the CAPM and the IAPM. Solnik (1974) then proposes a multifactor specification that combines the international and domestic indices.[12] This specification is consistent with the multifactor character of the linear factor model and the broader APT framework and the author argues that prices are dependent on international influences even though there is a high degree of dependence on domestic factors. The study represents an early consideration of international and domestic influences in return models. In later work, Solnik (1983: 452) lays down the theoretical foundation for an extension of the conventional APT framework to allow for international influences. This is done by demonstrating that the structure of the APT relation is invariant to the currency chosen by stipulating two other invariance propositions. The first is that the structure of the linear factor is invariant to the choice of currency. The second relates to the existence of an international risk free portfolio and stipulates that if returns are assumed to follow a homogeneous stochastic process, then any arbitrage portfolio that is riskless in a given currency will be riskless in any other currency (Solnik, 1983: 451, 453). For these invariance propositions to hold, exchange rates, similarly to returns, must be determined by the same *k*-factor linear factor model (Cho, Eun & Senbet, 1986: 315).

Cho *et al.* (1986: 313; 325) state that the invariance propositions laid down by Solnik (1983) permit the testability of the international APT framework. Using return data for 11 countries[13] and inter-battery factor analysis, it is reported that there are between one and five international factors that describe returns in the linear factor model across country pairs and that there are between three and four common international factors across all markets. Further tests indicate that the risk-free rate is the same between most country pairs but that the factor risk premia differ across countries and that both the risk premia and the risk-free rate differ between country groups. The authors state that while these findings do not provide support for the IAPT and international integration, they do not rule out the possibility of the APT being valid for local or regional markets.[14] While these results are somewhat

[12] The residuals of the domestic index are used in a regression of the domestic index onto the world index. This represents a purely national factor (Solnik, 1973: 374).

[13] Cho *et al.* (1986) use return data for stocks according to country membership. Countries in the sample are the US, Canada, France, Germany, the Netherlands, Switzerland, United Kingdom, Australia, Hong Kong, Singapore and Japan. Tests are conducted by combining two different countries in the sample.

[14] The authors report that all three hypotheses (of same risk-free rate, same risk premia, same risk-free rate and risk premia) are not rejected in 20 (22) in terms of the Dollar (Yen) out of the 55 country pairs considered.

ambiguous and not fully conclusive, the study suggests that non-domestic (at the very least, regional) common factors feature in the linear factor model and to some extent, in the APT relation. Korajczyk and Viallet (1989: 568, 589) undertake a study of international risk in four developed markets, namely the US, Japan, the UK and France. Although preliminary tests indicate that over 15 international factors are reflected in returns, the authors settle for a more practical structure of between five and 10 factors. Returns on domestic market indices are regressed onto the first five statistically derived factors and, with the exception of France, [15] all markets are found to be strongly related to these factors. The authors argue that these results point towards several common international influences in returns. Furthermore, a comparison of the performance of multifactor APT specifications to CAPM alternatives indicates that the APT outperforms a value-weighted version of the CAPM and that the APT outperforms the equal-weighted CAPM following adjustments for seasonalities, as evident from lower pricing errors.

The findings of Cho *et al.* (1986) and Korajczyk and Viallet (1989), building upon the extensions of Solnik (1974; 1983), indicate a role for non-domestic factors in multifactor descriptions of time series variation in returns and asset pricing. These studies, together with the theoretical foundations set out by Solnik (1973; 1983), represent an early extension of the APT that recognises that international influences may be important for explaining return behaviour and asset pricing. The role of international influences in stock returns and the APT is discussed further in Chapter 4.

2.3.3. Limitations

Support for the main propositions of the APT indicates that it is a general framework within which the multifactor structure of the return generating process can be explored and pricing relationships can be established. Outlined below are some of the main limitations and challenges to the APT, which have been noted in the literature. Importantly, one such limitation contributes to an important extension of the framework, which is of direct relevance to this study. This limitation, together with a number of others, is outlined below.

Dhrymes, Friend, Gultekin and Gultekin (1985: 661) show that not only does the number of factors increase with the number of stocks in a sample but also with the sample length. Using return data for stocks in the CRSP database, groups of 30, 60 and 90 stocks are

[15] The reported $\bar{R}^2$ s range between 0.31 (France, value-weighted) and 0.99 (UK, equal-weighted) (Korajczyk & Viallet, 1989: 568).

formed and the full sample is divided into two subperiods. The authors report that for each of the two subperiods, five factors are sufficient for groups of 30 stocks but eight factors are sufficient for groups of 60 stocks and 13 factors are sufficient for groups of 90 stocks. The number of factors for each respective group size increases to seven, 11 and 17 respectively when the full sample period is considered. In earlier work, Dhrymes, Friend and Gultekin(1984: 324, 340) state that there is no reason why the number of factors should not keep on increasing as the sample size increases. The instability in the number of factors and a lack of an established upper bound can be seen as a limitation of the APT framework, especially when attempting to describe the linear factor model.

While Roll and Ross (1980), Kryzanowski and To (1983), Chen (1983) and others find that factors derived from the linear factor model are priced in the expected returns, Dhrymes *et al.* (1982: 36) present evidence to the contrary. In a re-examination of the findings of Roll and Ross (1980) and using a similar sample, the authors find that the risk premia are jointly significant in under a seventh of the sample groups considered.[16] The significance of risk premia deteriorates further when standard deviation and skewness are included in cross-sectional tests of the APT. Risk premia are now jointly significant in under a 20th of the groups, indicating that factor loadings on common factors lose their explanatory power when considered alongside idiosyncratic (residual variance and skewness) factors. Dhrymes *et al.* (1982:36) argue that these results "suggest a very substantial failure for one of the crucial implications of the APT model."

The APT framework is touted as an alternative to the CAPM. For it to be a viable alternative, it should explain anomalies not explained by other alternatives. Tests of whether this is the case may be considered as alternative comparisons on criteria other than comparisons on the conventional measure of explanatory power, the $\bar{R}^2$. Examples of two such anomalies are the size and January effect. Gultekin and Gultekin (1987: 1221) investigate whether the January effect impacts tests of the APT model. Using CRSP return data, it is reported that the risk premia are always jointly statistically significant when January returns for groups of 30 and 90 stocks are used in cross-sectional regressions. However, when average returns for the remaining 11 months are considered and exclude January returns, the risk premia are jointly significant for under a 10th of the groups of 30 stocks and under a third of groups of 90 stocks. Gultekin and Gultekin (1987) conclude that the APT, much like the CAPM, can

[16] This corresponds to six out of 42 groups of 30 stocks each drawn from stocks listed on the NYSE and AMEX.

only explain the January risk-return relationship. This indicates that in the context of seasonal anomalies, the APT may not be a superior alternative to the CAPM. In a similar vein, Reinganum (1981: 315; 317) investigates whether the APT is able to account for the size-effect. Factor loadings are estimated using CRSP data and stocks are grouped into control portfolios with similar factor loadings. It argued that stocks with similar factor loadings should have similar average returns. Excess returns are then estimated for stocks using the control portfolios. As these are risk-adjusted returns, they can be used to test the APT's ability to explain the firm size effect. Stocks are then sorted into size-based portfolios on market value. As the excess returns are assumed to be risk-adjusted, average excess returns should jointly be equal to zero across portfolios formed on market value. Reinganum (1981)) reports that average excess returns are not jointly equal to zero for three, four and five factor models. Portfolios comprising large firms are characterised by the lowest returns and the difference between average excess returns on the portfolios of smallest and largest stocks is statistically significant. It is argued that this evidence is inconsistent with the APT and that the APT fails to account for anomalies that are unaccounted for by the CAPM. However, Reinganum (1981) emphasises that a number of hypothesis are jointly investigated and therefore it is difficult to identify the source of inconsistency in the APT. It is argued by the author that inconsistencies may arise from non-linearity of the return generating process, the inability to diversify away idiosyncratic risk, the absence of cross-sectional relationship and the existence of arbitrage opportunities over the sample period.

The final and perhaps most important limitation, which spurs further development of the APT, is that factors derived using factor analysis in early studies and the associated factor risk premia are unidentified and uninterpretable (Priestley, 1996: 870; Reilly & Brown, 2012: 243). Yli-Olli and Virtanen (1992: 510) acknowledge the limitations of factor analysis in the APT and state that the signs on factor coefficients have no meaning and the scaling of coefficients and associated risk premia is arbitrary. It is also difficult to determine the correct number of factors. This is aptly demonstrated in the early studies of the APT, which rely upon factor analysis. Dhrymes *et al.* (1982:1) aptly summarise the consequences of this limitation by stating that "without ascribing economic meaning to the factors on which the APT is based, it is difficult to see how the empirical implementation of the arbitrage pricing theory might be useful either for explanatory or predictive purposes." French (2017: 12) postulates that a weakness of empirical tests of the APT is that it does not specify the factors that will enter the APT. Chen (1983: 1409) recognises this limitation early on and states that the economic interpretation of common factors is "probably the most important direction for future research." It is suggested that the limitation of unidentified factors may be resolved in two ways. The first is to specify a theory that aids in the identification of factors that impact returns. The second is to examine returns and to determine to which macroeconomic factors they correspond.

2.3.4. The Macroeconomic APT

The next step in the development of the APT framework, motivated by the limitation of unidentifiable factors and uninterpretable risk premia, is the extension of the framework to the macroeconomic APT. Chen *et al.* (1986: 383-384) acknowledge that stock prices react to economic news but argue that theory has been silent on which events are likely to influence stock prices in a systematic manner. The authors state that "a rather embarrassing gap exists between the theoretically exclusive importance of systematic 'state variables' and our complete ignorance of their identity." (Chen *et al.,* 1986: 384) Consequently, the macroeconomic APT assumes that returns respond to macroeconomic shocks and relies upon observable macroeconomic factors to represent and measure the pervasive factors in returns (Connor, 1995: 42).

The first published study[17] to employ macroeconomic factors in place of statistical factors is that of Chan, Chen and Hsieh (1985: 452), who set out to study the size effect using stocks listed on the NYSE. The authors apply existing theory, namely the dividend discount model, to identify and to motivate for a set of macroeconomic factors that may impact stock prices (also see Birz & Lott, 2011: 2793). It is hypothesised that asset prices are related to the sum of discounted expected cash flows and therefore returns are impacted by changes in expected cash flows or the discount rate or both. These changes are related to changing macroeconomic conditions. Six macroeconomic factors are suggested; the growth in industrial production, unanticipated inflation, changes in expected inflation, the term structure, the change in net business formation and the risk premium. Returns on the value- and equally-weighted NYSE indices are also considered and fulfil the role of stock market proxies. The Fama-MacBeth (1973) approach to testing the APT relation and the firm-size effect is applied (Section 2.2.). To derive inputs, returns on 20 size-based portfolios are

[17] Chan *et al.*(1985) are the first to publish a paper that employs macroeconomic factors. However, the seminal work of Chen *et al.* (1986) is widely cited as laying the foundation for multifactor models employing macroeconomic factors. At the time that Chan *et al.*(1985) published their work, the seminal Chen *et al.* (1986) paper was an unpublished working paper.

regressed onto the macroeconomic factors in the linear factor model and factor betas are then used in the cross-sectional APT relation (Chan *et al.,* 1985: 456-457):

$$\begin{aligned} r_i &= \lambda_0 + \lambda_1 \hat{b}_i(EWNY) + \lambda_2 \hat{b}_i(IP1SA) + \lambda_3 \hat{b}_i(DEI) + \lambda_4 \hat{b}_i(UITB) + \lambda_5 \hat{b}_i(PREM) \\ &+ \lambda_6 \hat{b}_i(UTS) + \varepsilon_i \end{aligned} \tag{2.6}$$

where r_i are the returns on the 20 size based portfolios, λ_0 is the intercept, λ represents the respective risk premium associated with each exposure to a factor and the $\hat{b}_i$ s are the respective exposures associated with factors in the linear factor model, as denoted by $(IP1SA)$, (DEI), $(UITB)$, $(PREM)$ and (UTS), and the market proxy $(EWNY)$. Chen *et al.* (1985) consider the market aggregate to be a proxy for real activity and argue that it will reflect information in an efficient market. The residuals, ε_i, are interpreted as risk-adjusted returns in the study. Equation (2.6) emphasises the role of the underlying linear factor model which must be first formulated and estimated to obtain inputs, the $\hat{b}_i$ s, for testing in the APT relation. The risk premium $(PREM)$, industrial production $(IP1SA)$*,* and unexpected inflation $(UITB)$*,* are found to be priced during the overall period (1958-1977) whereas expected inflation $(DEI)$and the term structure (UTS) are priced during some of the subperiods,[18] together with the first three factors. As pre-specified macroeconomic factors are now used, an interpretation can be ascribed to the results. For example, Chan *et al*. (1985) state that the negative risk premium associated with the term structure implies that long-term bonds provide a hedge against shifts in the interest rate. Also, it is reported that the growth in net business formation, $(BUSF)$*,* is priced and competes with the risk premium, indicating that both factors are proxies for changing business conditions. In testing for the firm size effect, the authors report that mean differences in the residuals (risk adjusted returns) of cross-sectional regressions for portfolios formed on the basis of firm size, are jointly equally to zero. This implies that macroeconomic factors provide an explanation of the size effect, which is assumed to be a proxy for unidentified factors (Chan *et al.,* 1985: 464, 468). Chan *et al.'s* (1985) study is a useful early example of an application of the APT using macroeconomic factors in place of statistically derived factors and demonstrates how the use of pre-specified macroeconomic factors permits an interpretation of the results.

[18] There are 1958 to 1972, 1958 to1967 and 1968 to 1977.

Chen *et al.* (1986: 384) state that the aim of their study is an exploration of the underlying factors that influence stock returns. As in Chan *et al.* (1985), the authors rely upon the dividend discount model to inform the set of macroeconomic factors that is perceived to influence returns.[19] Accordingly, eight core macroeconomic factors are suggested together with two market indices, in the form of equally- and value-weighted indices comprising NYSE-listed stocks. Chen *et al.* (1986) argue that the market indices will timeously reflect public information relating to the macroeconomic state. A preliminary linear factor model (as opposed to the APT relation in equation (2.6.)) that is the basis of the study is defined (reproduced) as follows:

$$R = a + b_{MP}MP + b_{DEI}DEI + b_{UI}UI + b_{UPR}UPR + b_{UTS}UTS + e \tag{2.7}$$

where the betas in equation (2.7) are sensitivities to the respective macroeconomic factors that characterise the linear factor model, as before and MP is the growth rate in monthly industrial production, DEI is the change in expected inflation, UI is unexpected inflation, UPR is the risk premium and UTS is the term structure of interest rates. Equation (2.7) is a representation of the linear factor model characterised by pre-specified macroeconomic factors in place of unidentified statistically derived factors. Using returns on 20 equally-weighted portfolios, Chen *et al.* (1986) estimate equation (2.7) as a time series model and use the resultant betas in cross-sectional tests of pricing relationships. Changes in industrial production and the risk premium are found to be consistently important in explaining expected returns. Other factors that also feature, but are less important, are the inflation factors, DEI and UI, and the term structure, UTS. The inclusion of the value and equally-weighted indices, in addition to the factors in equation (2.7) does not impact the initial pricing results indicating that these factors reflect priced risk in excess of that reflected by the respective market betas. The authors conclude that stock returns are exposed to systematic economic news, macroeconomic factors are priced in accordance with their exposures and that the identification of relevant macroeconomic factors can be accomplished by reference to financial theory (as initially suggested by Chen, 1983). Furthermore, by specifying and

[19] To show that the macroeconomic factors considered are proxies for pervasive influences, Chen *et al.* (1986) extract five factors from returns in the sample and then regress factor scores onto the macroeconomic factors. A macroeconomic factor is confirmed as being related to stock price movements if it is significantly related to at least one of the statistically derived factors. Production growth, term structure and the risk premium are related to these factors but support for the inflation factors is weak (Chen *et al.,* 1986: footnote 7).

acknowledging the linear factor model, Chen *et al*. (1986) acknowledge a linear factor model characterised by macroeconomic factors as an underpinning of the APT.

Hamao (1988: 47; 51-52) extends the application of the macroeconomic APT to the Japanese stock market. The same factors as in Chen *et al.* (1986) are considered but the factor set is extended to include the exchange rate (Yen/Dollar), the oil price and the terms of trade. This is motivated by the openness of the Japanese economy and dependence on international trade. Value- and equally-weighted indices comprising stocks on the Tokyo Stock Exchange (TSE) are also considered. These are again treated as proxies for numerous economic shocks and aggregators of information. Notably, Hamao (1988) acknowledges the basic linear factor model (which includes the core factors) and the APT relation by stating the respective specifications (reproduced) and follows the by now familiar Fama and Macbeth (1973) approach:

$$R_p = a_p + b_{1p}MPSA + b_{2p}DEI + b_{3p}UI + b_{4p}UPR + b_{5p}UTS + \varepsilon_p \tag{2.8}$$

$$R_p = \lambda_0 + \lambda_1 b_{1p} + \lambda_2 b_{2p} + \lambda_3 b_{3p} + \lambda_4 b_{4p} + \lambda_5 b_{5p} + \varepsilon_p \tag{2.9}$$

where equation (2.8) is the linear factor model and equation (2.9) is the cross-sectional APT relation. The link between the APT model and the linear factor model and the underpinning role of the linear factor model in the overall APT framework is immediately apparent; the betas in equation (2.8) are the explanatory factors in equation (2.9). For this reason, the adequacy of the specification (or lack thereof) of the linear factor model is important and of concern. In equation (2.8), $MPSA$ is industrial production, DEI are the changes in expected inflation, UI is unanticipated inflation, UPR is the default spread and UTS is the term structure. Both equations are estimated using returns on 20 portfolios comprising Japanese stocks. Hamao (1988) reports that the growth rate in industrial production, changes in expected inflation, unanticipated inflation and unanticipated changes in the risk premium and the term structure are (generally) priced. Further tests show that the terms of trade, exchange rates and oil prices are not priced. It is suggested that this is because the pricing influence of these factors is reflected in other factors. As in Chen *et al.* (1986), the equally- and value-weighted TSE indices are not priced, indicating that these indices do not capture additional systematic risk in the cross-sectional APT model. Hamao (1988) explicitly recognises the underlying relationship between the linear factor model and the APT and shows that the macroeconomic APT is applicable and relevant in a non-US market. It is also shown that factors other than market betas are relevant for pricing. Implicitly, this is a conceptual confirmation of a multifactor linear factor model; the macroeconomic factors that feature in the linear factor model are also relevant in a pricing context.

The studies discussed above present an evolution of the APT. By referring to financial theory in the form of the dividend discount model, macroeconomic factors are identified and are postulated to be proxies for pervasive influences in returns. Evidence of pricing indicates that this approach is consistent with the central propositions of the APT. The linear factor model is now characterised by macroeconomic factors and consequently, the risk premia are interpretable. Moreover, these studies mark the beginning of the widespread application of the macroeconomic APT framework across global markets.

2.4. FURTHER APPLICATION

The macroeconomic APT is a major milestone in the development of the APT and continues to be applied in empirical analysis. The studies cited below offer a limited but hopefully informative overview of its application and flexibility, which at times departs from its traditional asset pricing focus. All studies cited below refer directly to the APT as motivation.

Berry *et al.* (1988: 29-31) apply the APT to model the return generating process of seven economic sectors and 82 industrial sectors comprising the US stock market. Five factors are used to describe the linear factor model, namely changes in the default premium, changes in the term structure, unanticipated inflation/deflation, changes in real final sales and the residual market factor derived from returns on the S&P 500 Index. [20] The residual market factor is treated as a proxy for omitted factors.[21] The economic groups and industrial sectors in the sample are broadly responsive to innovations in these factors and the average respective $\bar{R}^2$ s are 0.611 and 0.485. The widespread observed significance of these factors and the observed explanatory power indicates that innovations in these factors can explain the return generating process. Berry *et al.* (1988) further ascribe an interpretation to the results by comparing risk profiles (exposures to the respective factors) across economic groupings. The authors also argue that on the basis of these findings, a risk sterilisation strategy may be formulated and propose that the APT-motivated linear factor model specification can be applied to formulate an investment strategy that outperforms a

[20] The choice of these factors is motivated by the work of Chan *et al.* (1985) and Chen *et al.* (1986).

[21] Chapter 3 provides a comprehensive treatment of the residual market factor and its role in the APT framework.

benchmark. Importantly, this study departs from the early pricing orientated APT studies and instead demonstrates an application to the modelling of the return generating process.

Beenstock and Chan (1988: 34) apply the APT to the UK stock market but consider a different and a broader set of pre-specified macroeconomic factors than that of Chan *et al.* (1985), Chen *et al.* (1986) and Hamao (1988).[22] Using return data extracted from the London Share Price Database (LSPD), the linear factor model is found to be described by four factors, namely the UK treasury bill rate, the broad money supply, fuel and material input costs to manufacturing and the UK retail price index. The use of these factors, and the consideration of a broader set of factors, presents a departure from the conventional factors of industrial production, changes in expected inflation, unanticipated inflation, the default spread and the term structure used in the early seminal studies outlined in Section 2.3.4. Using returns on 76 portfolios, Beenstock and Chan (1988) report that the four factors that describe the linear factor model are also priced and explain approximately 30% of the variation in expected returns.[23] The authors conclude that the contribution of the study lies in the avoidance of the use of factor analysis and the specification of a return generating process characterised by macroeconomic factors. The study also contributes by applying the APT to yet another market and showing that relevant factors may differ across markets. In contrast to Beenstock and Chan (1988), Poon and Taylor (1991: 620, 630) investigate whether the conventional Chen *et al.* (1986) factors *are* applicable to the UK stock market. Value- and equally-weighted market indices are also included in the factor set. Univariate pricing tests indicate the absence of significant contemporaneous pricing relationships between these factors and returns on portfolios formed from stocks in the LSDP. The significance of lead and lag relationships is also tested and the results indicate that the equally- and value-weighted indices, the term structure, growth rate in monthly industrial production and unanticipated inflation are priced at differing lags. Poon and Taylor (1991) state that pricing relationships may not be contemporaneous and that it has been shown, that at the very least, the conventional factors are not relevant for pricing in the UK stock market. These findings, taken together with those of Beenstock and Chan (1988), indicate that the structure of the linear factor model and the APT relation differ across markets. Also,

---

[22] The other factors that are considered are the UK general index of wages, industrial stoppages, an export volume index, a retail volume index, relative export prices, GDP and total OECD production.

[23] The results vary slightly according to methodological variations used in the estimation of the linear factor model.

these results propose a further avenue of research, namely that of establishing which factors are relevant as opposed to determining whether a particular set of factors is relevant.

Similarly, to Beenstock and Chan (1988) and Poon and Taylor (1991), Clare and Thomas (1994: 310) also consider a large factor set but investigate a further aspect of the APT in the UK stock market, namely the impact of portfolio formation on results. Portfolios of UK stocks are formed on the basis of market beta and market value. Notably a large and alternative set of 20 macroeconomic factors is considered. This is motivated by the need for the factor set to reflect the "small, open-economy nature of the UK" (Clare & Thomas, 1994: 310). Results indicate that when portfolios are formed on the basis of beta, six factors are priced, namely oil prices, debentures and loan redemption yields, the default spread, the comfort index,[24] the retail price index, private sector lending and the current account balance. However, when portfolios formed on the basis of market value are used in testing pricing relationships, the results are conspicuously different; only the comfort index and the retail price index are priced. Clare and Thomas (1994) comment that this result suggests that pricing is not invariant to the method employed in constructing portfolios. The contribution of this study lies in further interrogating the APT framework by showing that portfolio formation may impact the results of the APT model, hitherto an unexplored aspect. Also, the authors explicitly acknowledge that the factor set should consider the nature of the economy. [25] This study demonstrates the plethora of avenues of research that stem from the implementation of the APT and that are concerned with the identity of priced factors.

Chen and Jordan (1993: 66) compare the ability of the statistical APT and the macroeconomic APT relation to explain returns using industry portfolios formed from firms in the CRSP database and also test whether macroeconomic factors are proxies for pervasive influences. The Chen *et al.* (1986) factors comprise the macroeconomic factor set and the unexpected changes in oil prices and returns on a market index, a value-weighted portfolio of NYSE-listed stocks, are also considered. The first set of results relates derived factor scores to the seven pre-specified factors. The results indicate that the market index, changes in the term structure, changes in the risk premium, and changes in the oil price are related to these five statistical factors and that changes in industrial production are weakly related to these factors. The most significant finding is that the macroeconomic factors considered appear to be partial proxies for the pervasive influences in returns represented

[24] Defined as the ratio of the consol to equity market dividend yield (Clare & Thomas, 1994: 311).
[25] With the caveat that this is the broadest set of factors considered at the time of publication of this study.

by the statistical factors; Chen and Jordan (1993: 74) report that the $\bar{R}^2$s range between 0.495 for the first factor and 0.022 for the fifth factor and average 0.173 (also see Connor, 1995: 41). Cross-sectional tests indicate that the statistical APT outperforms the macroeconomic APT in explaining expected returns; the respective $\bar{R}^2$s are 0.374 and 0.314. Of the pre-specified factors, the market index, changes in expected inflation and unanticipated changes in oil prices are priced. Chen and Jordan's (1993) approach is significant in that it relates statistical factors to pre-specified factors. This is one of the solutions proposed by Chen (1983) to identifying macroeconomic factors. The study also proposes that a statistical APT outperforms a macroeconomic APT in explaining the cross-section of returns.

Arguing for an application of the APT to the real estate sector, Chen, Hsieh and Jordan (1997: 505) state that "if more than one variable does play a significant role in real estate returns, then the arbitrage pricing theory (APT) proposed by Ross (1976) would seem to be a natural selection as a theoretical framework for studying the real estate return generating function" (also see Chan, Hendershott & Sanders, 1990; Chen *et al.,* 1998). Consequently, the APT framework is applied to investigate the impact of the five macroeconomic factors on equity real estate investment trust (EREIT) returns in the CRSP database and is compared to a statistical linear factor model. The macroeconomic factors are the changes in the term structure, changes in the risk premium, expected inflation, changes in expected inflation, unexpected inflation and returns on a residual market factor (derived from the CRSP value-weighted index). The macroeconomic linear factor model significantly outperforms the statistical linear factor model in terms of explanatory power, as measured by average $\bar{R}^2$s, for two out of the three subperiods considered and for the full sample period.[26] The Davidson and MacKinnon (1981) test confirms that the macroeconomic linear factor is superior relative to the statistical factor model. Cross-sectional tests show that unanticipated inflation, the unanticipated change in the term structure, the residual market factor and the default spread are priced although pricing differs across subperiods. Nevertheless, for two out of three subperiods, more than one factor is priced. Chen *et al.* (1997) conclude that the macroeconomic APT appears to be superior and that the study provides further guidance for applications of the APT to real estate related assets. Importantly, the study indicates that macroeconomic linear factor models are a worthy

[26] These are the January 1980 to December 1985 and the January 1986 to December 1991 subperiods. The full sample period spans the period January 1974 to December 1991.

competitor to statistical linear factor models, although this may not be the case when comparing the cross-sectional explanatory power of the statistical and macroeconomic versions of the APT, as in Chen and Jordan (1993).

Elton *et al*. (1995: 1230, 1232) state that although the US bond market is far larger in market value than the equity market, it has not been studied within the context of the APT. A total of six factors are proposed to explain bond returns, these being common-stock index returns (the S&P 500 Index), the default spread, a mortgage return series to capture option elements, an aggregate bond index (index factors), changes in the Gross National Product (GNP) and changes in expected inflation (termed as fundamental (macroeconomic) factors in the study). The time series $\bar{R}^2$ of the (initial) six-factor model for returns on passive bond portfolios ranges between 0.81 and 0.98 and averages 0.923. In the cross-section, the APT explains 82.47% of variation (Elton *et al.,* 1995: 1244-1245). Both macroeconomic factors, the unexpected changes in the real GNP and unexpected changes in inflation, are priced. The six-factor model that incorporates both indices and macroeconomic factors is compared to alternatives that include only the indices or a restricted number of indices. The authors report that both the indices and macroeconomic factors are important in explaining the time series behaviour of bond returns and that the inclusion of fundamental factors results in a large improvement in the explanation of expected returns. The study is an extension of the APT to an asset class other than equity for explanatory and pricing purposes . The findings of the study are that the linear factor model and the APT relation can be applied to describe and study the behaviour non-equity assets.

Antoniou, Garrett and Priestley (1998: 222) argue that assessments of the APT should show that the same factors price different subsets of stocks and carry the same prices of risk. [27] The authors investigate the generalisability of a return generating process on the London Stock Exchange by examining the relationships between macroeconomic factors for two samples, namely the estimation sample and a validation sample.[28] Using the estimation sample, a six-factor APT relation is arrived at with unexpected inflation, changes in expected inflation, the money supply, default risk, the exchange rate and excess returns on the market portfolio found to be priced. The factors identified in the estimation sample are those used

[27] Antoniou *et al.* (1998: 224) elaborate upon this by stating that there is no guarantee that a factor for one portfolio is also a factor for another portfolio – a factor is not common.
[28] Returns on a total of 138 stocks are used. These are randomly divided into two samples. The first sample is the estimation sample and the second sample is the validation sample.

as potential sources of risk in the validation sample. All factors with the exception of expected inflation are priced although default risk and the exchange rate have different risk premia. However, three of the risk premia, namely those associated with unexpected inflation, the money supply and returns on the market portfolio, have the same sign and are of a similar magnitude as in the estimation sample. Antoniou *et al.* (1998) comment that this demonstrates that the same factors can be used to price assets, suggesting that three factors are common to both samples. The explanatory power of the priced factors for both samples is compared and the specification is shown to explain a similar amount of cross-sectional variation in expected returns. These results are notable in that they show that macroeconomic factors are proxies for commo*n* factors that drive returns. This finding is analogous to those of early studies that rely upon factor analysis, notably those of Brown and Weinstein (1983) and Hughes (1984) (discussed in Section 2.3.1.). Antoniou *et al.* (1998) conclude by stating that it is possible to develop a unique return generating process that will explain a substantial amount of cross-sectional variation in expected returns across subsets.

Panetta (2002: 418) investigates the stability of the return generating process for the Italian stock market. It is acknowledged that the motivation is the non-equilibrium aspect of the APT represented by the linear factor model, namely that of relating returns to changes in the macroeconomic environment. Panetta (2002) refers to the dividend discount model to identify a broad set of relevant factors and similarly to Chen and Jordan (1993), applies factor analysis to reduce the factor set for investigating stability by relating factor scores to subsets of macroeconomic factors. A five-factor model that incorporates innovations in the term structure, unexpected changes in industrial production, unexpected inflation, unexpected changes in the oil price and changes in the Italian Lira/US Dollar exchange rate is identified. The model is estimated over non-overlapping periods by regressing returns on stocks listed on the Milan Stock Exchange onto these five factors. Results indicate that the relationships between returns and macroeconomic factors are generally unstable; the direction of impact is unstable, changing between successive periods and in magnitude (Panetta, 2002: 439-440).[29] It is proposed by the author that the observed instability may arise from cyclical economic variations during which the responses to economic fundamentals change and may also be attributable to the process of globalisation which may

[29] The number of sign reversals (sign changes) ranges between 14.1% (between period 3 and 4 for the term structure) to 71.7% (between period 1 and 2 for (also) the term structure). Sign reversals for other factors fall within this range.

modify exposure to economic shocks. Panetta's (2002) contribution to the literature lies in focusing on aspects other than pricing within the APT and the modelling of the return generating process. Instead, insight into the changing nature of the Italian economy is provided by applying the APT as a theoretical basis.

Cauchie *et al.* (2004: 168; 179) apply the APT to study the determinants of stock returns in the Swiss stock market, which is seen as being internationally orientated. The authors argue that most markets are imperfectly integrated and therefore returns may be determined by a combination of local and global risk factors. Four factors that are relevant to the Swiss economy are identified using cluster analysis. Two of these are international and two domestic, namely industrial production and expected inflation in G7 countries (global) and the Swiss term structure and the return on the Swiss stock market (local). According to Cauchie *et al.* (2004), this provides support for the hypothesis of a partially integrated stock market. However, none of the factors are found to be priced. Therefore, to establish a relationship between the macroeconomic factor risk premia and statistical factor risk premia, all of which are significant when negative and positive realisations are considered separately, the statistical risk premia[30] are regressed onto the macroeconomic risk premia and correlations are also established. The results indicate that each statistical factor risk premium is related to at least one macroeconomic risk premium. [31] Cauchie *et al.* (2004) show that statistical factor risk premia are associated with macroeconomic risk premia, suggesting that macroeconomic factors feature in expected returns. The study also demonstrates, through generalisation and extension, how the APT may be applied to study the role of international influences and market integration (Section 2.3.2.).[32]

Azeez and Yonezawa (2006: 569) investigate the Japanese stock market crash of the 1990s and argue that this event is attributable to fundamental changes in the Japanese economy that began in the 1980s. The study considers three periods, the pre-bubble period (1973-1979), the bubble economy (1980 to 1989) and the post-bubble period (1990-1998). The factors identified as important to the Japanese economy are the unanticipated shocks to the money supply, inflation, industrial production, exchange rates and land prices (Azeez & Yonezawa, 2006: 577). Results indicate that the money supply, inflation, industrial

[30] This is subject to the direction of factor realisations. Cauchie *et al*. (2004: 178) confirm that all factors are priced when risk premia are separated into positive and negative occurrences.

[31] For example, the correlation between the risk premia for Factor 1 and industrial production for the G7 countries is 0.49 (Cauchie *et al.,* 2004: 182).

[32] The international APT and international influences are discussed in detail in Chapter 4.

production, the term structure, the exchange rate and land are priced in the pre-bubble period whereas all of these factors, with the exception of the term structure are priced during the bubble and post-bubble periods. Risk premia are found to increase substantially (in absolute terms) during the bubble and post-bubble periods and the authors attribute this to the increase in risk during the crash period and thereafter. The authors go onto note that the factors that continue to be pervasive are associated with money supply factors, indicating that the growth of the money supply, attributable to low interest rates, is the most important factor driving Japan's stock prices. Azeez and Yonezawa's (2006) application of the APT complements that of Panetta (2002). The authors study the changing nature of risk and quantify risk related to an important economic event (the Japanese stock market crash) within the APT relation. In contrast, Panetta (2002) considers the stability of the return generating process within the APT framework.

Sadorsky (2008: 3855) relaxes some of the assumptions of the APT in an investigation of the impact of oil prices on US firms of different sizes. The impact of oil prices is studied within a multifactor model, which the author explicitly justifies with reference to the multifactor APT and cites the work of Ross (1976) as a theoretical basis (also see Faff & Chan, 1998: 21). Four economy-wide factors enter the model; market returns, interest rate spreads (term structure), oil prices and the oil price volatility. A firm size factor (firm sales) is also incorporated into the model and the impact of oil price changes is permitted to be asymmetric. This approach combines systematic factors with a firm-specific factor, firm size, presenting a departure from a "purist" APT implementation.[33] Another departure lies in the inclusion of oil price volatility to capture the asymmetric (non-linear) impact of oil prices. Sadorsky (2008) reports that larger firms, on average, have lower returns and that market returns, term structure, oil prices and volatility also have a significant impact on returns. The initial model specification is further adapted to allow for the asymmetric impact of oil prices and for interactions between oil prices and size. Results indicate that oil prices have an asymmetric impact on stock prices and firm size moderates the relationship between returns and oil price volatility (Sadorsky, 2008: 3858). On the basis of these results, the author outlines suggestions for policy makers as to how firms may be assisted when oil prices increase. Sadorsky's (2008) application of the APT emphasises its role as a general basis

[33] It can, however, be argued that the manner in which firm size is measured, namely in the form of firm sales, reflects macroeconomic conditions.

for a multitude of applications of multifactor specifications, even if these do not conform closely to the central propositions of the APT.

Szczygielski and Chipeta (2015: 2, 12) study the return generating process underlying the South African stock market, as represented by the JSE All Share Index, using a broad set of domestic factors and an international market index. Similarly to Sadorsky (2008), the motivation for the structure of the return generating process is the multifactor APT. The authors closely follow the prescripts of the APT and unexpected components are used in constructing factors. With reference to the dividend discount model, the linear factor model is characterised by changes in the inflation rate, inflation expectations, building plans passed (real activity), growth in the money supply, changes in the oil price, fluctuations in the exchange rate and innovations in the business cycle. Szczygielski and Chipeta (2015) also include the FTSE All World Index to account for unspecified global influences. All factors are statistically significant and the unrestricted model combining all of these factors explains 56% of the variation in returns. In further analysis, the influence of factors is decomposed by estimating restricted versions of specifications and returns on the JSE All Share Index are found to be mostly driven by movements of the FTSE All World Index. This is attributed to the high levels of integration of the South African stock market with global markets. Szczygielski and Chipeta's (2015) study is an example of an investigation of the return generating process framed within the APT.

The studies cited above demonstrate the ongoing application of the APT with asset pricing as the general orientation. However, these studies consider different aspects; the pricing of specific factors sets, different macroeconomic factor sets, the impact of portfolio formation, bond pricing, pricing in a specific sector, market integration and changing economic conditions. A number of studies, notably those of Berry *et al.* (1988), Panetta (2002) and Szczygielski and Chipeta (2015), focus on the return generating process. Sadorsky (2008) relaxes some the assumptions that underpin the APT and demonstrates how the APT is a theoretical basis for multifactor models in general. Though the studies cited differ in focus, they share the theoretical underpinnings of the APT and central to these applications is the underlying linear factor model. The linear factor model is required to either derive inputs for use in asset pricing tests or is of direct interest in itself. Additionally, the residual market factor is formally introduced in this section in the discussion of Berry *et al.'s* (1988) study of the return generating process for US economic and industrial sectors. This follows on from the mention of the role of market indices in Section 2.3.4. in Chan *et al.* (1985), Chen *et al.*

(1986) and Hamao (1988). In these studies, market indices are seen as aggregators of and proxies for economic information and additional pervasive influences.

2.5. CHAPTER SUMMARY AND CONCLUSION

The APT and the underpinning multifactor linear model is a response to the challenges to the CAPM and the underlying single-factor characterisation of the return generating process. The APT provides a multifactor framework for describing the return generating process of stock returns and for relating returns to risk in equilibrium. Aside from the structure of the linear factor model being of interest in itself, the linear factor model is a central component of the APT. Its formulation and estimation are required to derive inputs for the APT relation in the form of factor betas (Section 2.2.).

Early studies support the central tenets of the APT. These studies show that multiple factors characterise the return generating process, that the factors that feature in the linear factor model are priced, and that the factors that drive returns are systematic in nature (Section 2.3.1.). Early literature also introduces an extension to the APT in the form of the international APT. This attests to the general and flexible nature of this framework (Section 2.3.2.). The APT, as any asset pricing theory, is not without its limitations. The most important limitation in early studies is the use of statistically derived factors that are uninterpretable. This widely recognised limitation spurs the development of the APT and the response to this limitation is the macroeconomic APT (Section 2.3.3.). The macroeconomic APT relies upon pre-specified macroeconomic factors that are perceived to represent previously unidentified pervasive influences in stock returns. Chan *et al.* (1985), Chen *et al.* (1986) and Hamao (1988) provide support for the use of macroeconomic factors in the APT. In these studies, macroeconomic factors are priced in the APT relation suggesting that they are representative of the pervasive influences in stock returns (Section 2.3.4.). Following these seminal studies, it is shown that the macroeconomic APT is widely applied to numerous markets to investigate pricing and to describe the return generating process in the broader sense (Section 2.4.). It is also shown that the application of the APT is associated with numerous extensions. Common to both equilibrium and non-equilibrium applications is the underlying linear factor model. It is for this reason that the consequences of factor omission in the linear factor model are worthy of further consideration.

Two themes emerge from the literature discussed in this chapter. The first is that the linear factor model underpins all the studies discussed. For this reason, it is important to investigate whether the linear factor model can be adequately specified in terms of macroeconomic factors. This study investigates whether this is the case. The second theme follows from this and relates to the role of a market index in such studies. Consequently, the role of the residual market factor as a proxy for omitted pervasive influences in returns is investigated in this study. The role of the market index, and specifically the residual market factor in the APT, is discussed in the following chapter.

CHAPTER 3

THE MARKET INDEX, THE RESIDUAL MARKET FACTOR AND THE ARBITRAGE PRICING THEORY

3.1. INTRODUCTION

The emergence of the market index as a significant factor for stock returns is attributable to Sharpe (1963: 281), who proposes the diagonal model as a solution to the portfolio analysis problem.[34] The diagonal model assumes that returns are related to a common factor, which is the single most important influence on returns and therefore the diagonality assumption holds ($E(\varepsilon_{it}, \varepsilon_{it}) = 0$ (equation (2.2); Frankfurter & Lamoureux, 1990: 854)). Sharpe (1963) argues that amongst a number of factors, this single factor may be the level of the stock market as a whole. The market index has consequently come to be seen as the single most important factor influencing returns and has featured prominently in the diagonal model's successor, the market model. Fama and Macbeth (1973: 634) aptly articulate the role of the market index in the market model by stating that the market index is assumed to capture the effects of market-wide risks. Although the APT framework presents a departure from the assumption that the market index reflects all systematic risk, it nevertheless continues to feature in the APT. Specifically, the market index features either in its original form or as an orthogonalised (residualised) derivative in the form of a residual market factor.[35] The residual market factor is assumed to represent the influences of omitted and unobserved macroeconomic and psychological factors in linear factor models (Spronk & Hallerbach, 1997: 123; Deetz, Poddig, Sidorovitch & Varmaz, 2009: 299). APT literature treats the residual market factor as a solution to factor omission and the consequent underspecification of the linear factor model.

This chapter discusses the residual market factor by first outlining seminal work, specifically that of Burmeister and Wall (1986), which introduces the residual market factor as a concept

[34] Defined as the 1) formulation of probabilistic estimates of the future performance of assets, 2) the analysis of these estimates to determine an efficient set of portfolios and 3) the selection of portfolios best suited to an investor's preferences (Sharpe, 1963: 277).

[35] The reader's attention is drawn to the use of the terms the *market index* and the *residual market factor.* Some parts of this chapter will discuss literature dealing with the market index. However, inferences drawn from this literature are easily generalisable to the residual market factor as the residual market factor is directly derived from a market index. Therefore, the residual market factor is a close approximation of the market index but with a desirable statistical property, namely the lack of correlation with factors included in a specification.

and then setting out its role in the APT framework. Consideration is also given to the nature of the information that enters the market index, and by extension, the residual market factor.

The rest of this chapter proceeds as follows; Section 3.2. summarises the seminal work of Burmeister and Wall (1986), who are the first to apply the residual market factor and lays down the theoretical foundations of the residual market factor. Section 3.3. argues for an approach favouring the use of the residual market factor over other alternatives to resolve factor omission bias. Section 3.4. discusses the use of the residual market factor in the APT literature and Section 3.5. provides insight into the informational content of market indices, and by implication, the residual market factor. Section 3.6. concludes and summarises this chapter.

3.2. INTRODUCING THE RESIDUAL MARKET FACTOR

Burmeister and Wall (1986: 3, 9) introduce the residual market factor in an early study of the return generating process contextualised in terms of the APT that seeks to model the linkages between returns on the S&P 500 Index, a portfolio of randomly selected stocks, a fund (the T.Rowe Price New Horizons Fund) and the macroeconomic environment.[36] Owing to this paper's importance, the propositions and findings of this paper are briefly summarised in the discussion that follows.

To explain returns, innovations (unanticipated changes) in the default spread, the term structure, inflation and real final sales are used in a preliminary restricted multifactor specification. Burmeister and Wall (1986) argue that these macroeconomic factors can be seen as being representative of the pervasive influences (that would otherwise be reflected in statistical factors), which impact returns by influencing expected cash flows and/or the discount rate. An excerpt of the results is reported below (Burmeister & Wall, 1986: 8):

[36] Burmeister and Wall's (1986) sample consists of total returns on an equally-weighted portfolio of 20 randomly selected stocks and total returns on the T.Rowe Price New Horizons Fund for the December 1971 to November 1981 period.

Table 3.1: Time Series Regression Results (1)

r_t	Const.	$UPR(t)$	$UTS(t)$	$UI(t)$	$UGS(t)$	$\bar{R}^2$
1) S&P 500 total returns index	0.0094 (2.64)	1.54 (4.57)	0.50 (4.39)	-3.03 (-2.64)	1.30 (4.38)	0.29
2) Total Return on equally weighted portfolio of 20 randomly selected stocks.	0.0110 (2.43)	2.19 (5.02)	0.58 (4.00)	-4.20 (-2.82)	1.60 (4.15)	0.29
3) Total Return on T. Rowe Price New Horizons Fund	0.0120 (2.08)	1.74 (3.30)	0.49 (2.75)	-5.78 (-3.22)	1.89 (4.07)	0.22

Notes:
t-statistics are reported in parentheses as in Burmeister and Wall (1986)
Monthly sample period from December 1971 to November 1981.
$UPR(t)$ = unanticipated change in the risk premium, $UTS(t)$ = unanticipated change in the term structure, $UI(t)$ = unexpected inflation, $UGS(t)$ = unexpected growth in real sales

Source: Burmeister & Wall (1986)

By relating returns on the S&P 500 Index (row (1)), a portfolio of 20 randomly selected stocks (row (2)) and the T. Rowe Price New Horizons Fund (row (3)) to the four factors in Table 3.1., Burmeister and Wall (1986) present a description of the return generating process motivated by the linear factor model. The model estimated in row (1) is of particular importance; the residuals of this regression are used in subsequent unrestricted versions of the models in row (2) and (3). These residuals are the residual market factor; the innovations in the market index that are not explained by the first four macroeconomic factors (Burmeister & Wall, 1986: 9). The results of the unrestricted model that incorporates the residual market factor are reported in Table 3.2. (Burmeister & Wall, 1986: 8):

Table 3.2: Time Series Regression Results (2)

r_t	Const.	$UPR(t)$	$UTS(t)$	$UI(t)$	$UGS(t)$	$UM(t)$	$\bar{R}^2$
4) Same as row (2)	0.0110 (4.76)	2.19 (9.84)	0.58 (7.75)	-4.20 (-5.53)	1.60 (8.13)	1.12 (18.10)	0.82
5) Same as row (3)	0.0120 (3.90)	1.74 (6.20)	0.49 (5.16)	-5.78 (-6.05)	1.89 (7.64)	1.32 (17.10)	0.78

Notes:
t-statistics are reported in parentheses as in Burmeister and Wall (1986)
Monthly sample period from December 1971 to November 1981.
$UPR(t)$ = unanticipated change in the risk premium, $UTS(t)$ = unanticipated change in the term structure, $UI(t)$ = unexpected inflation, $UGS(t)$ - unexpected growth in real sales $UM(t)$ = changes in the market index not explained by $UPR(t)$, $UTS(t)$, $UI(t)$ and $UGS(t)$ (notation reproduced).

Source: Burmeister & Wall (1986)

Burmeister and Wall (1986) note that the increase in the adjusted coefficient of determination, $\bar{R}^2$, in rows (4) and (5) is what is to be expected with the inclusion of the S&P 500 Index. However, an additional four distinct risk types also feature in the linear factor model, aside from market risk which is measured by the residual market factor, $UM(t)$. According to the authors, rows (4) and (5) provide an alternative interpretation of the APT approach to modelling the return generating process. Although much (or even most) of the variation in returns can be explained by the market model, the approach permits a breakdown of total market risk attributable to four systematic factors and to what Burmeister and Wall (1986: 10) term as "other market risk." To permit these factors to exert an impact on returns, the residual market factor is derived from the residuals of an auxiliary regression of a market index onto the remaining macroeconomic factors in the linear factor model. Its purpose is to control for omitted and unobservable factors, and, as the residual market factor is uncorrelated by construction with the remaining factors, it has no impact on the coefficients on the remaining factors (Deetz *et al.,* 2009: 299; Czaja *at al.,* 2010: 130).

With reference to notation, the residual market factor may be defined as the portion of returns on the market index, R_{Mt}, which is not explained by the factors incorporated into a linear factor model:

$$R_{Mt} = E(R_i) + \sum_{k=1}^{K} b_{ik} f_{kt} + M\varepsilon_t \tag{3.1}$$

where R_{Mt} is the return on a broad market index and $\sum_{k=1}^{K} b_{ik} f_{kt}$ is a set of factors with the associated sensitivities that features in the general linear factor model applied across assets. The residuals of this regression, $M\varepsilon_t$, constitute the residual market factor and are included in subsequent specifications of the linear factor model across assets in the sample:

$$R_{it} = E(R_i) + \sum_{k=1}^{K} b_{ik} f_{kt} + b_{iM\varepsilon} M\varepsilon_t + \varepsilon_{it} \tag{3.2}$$

where $M\varepsilon_t$ in equation (3.2) is the residual series from equation (3.1) and the sensitivity of R_{it} to $M\varepsilon_t$ is denoted by $b_{iM\varepsilon}$. By construction, the residual market factor is uncorrelated with the orthogonalising set of factors. Van Rensburg (1996: 107), in a study of APT factors in the Johannesburg Stock Exchange (JSE), demonstrates that this is indeed the case (also see Wurm & Fisicaro, 2014). The residual market factor is constructed by regressing returns on the JSE Actuaries All Share Index onto four factors that are hypothesised to have a pervasive impact on returns.[37] These factors are innovations in the Rand gold price, the Dow Jones Industrial Index (DJIA), inflation expectations and the term structure of interest rates. The correlations between the residual market factor (denoted by UM in the study), the returns on the conventional market index (R_m), the JSE Actuaries All Share Index and the four factors that are used in the construction of residual market factor are reproduced faithfully from Van Rensburg (1996: 107):

Table 3.3: Correlation Matrix: Macrovariables

	R_m	*UGOLD*	*UDJ*	*UINF*	*UTSD*
UGOLD	0.35	-			
UDJ	0.30	0.02	-		
UINF	-0.37	-0.32	-0.07	-	
UTSD	0.05	0.11	-0.03	-0.65	-
UM	0.83	0.00	0.00	0.00	0.00

Notes:
UGOLD = unanticipated changes in the Rand gold price, *UDJ* = unanticipated returns on the Dow Jones Industrial Index (DJIA), *UINF* = unanticipated changes in inflation expectations, *UTSD* = unanticipated changes in the term structure of interest rates, R_m = returns on the JSE Actuaries All Share Index, *UM* = the residual market factor

Source: Van Rensburg (1996)

As evident from Table 3.3., the residual market factor, UM, is uncorrelated with the remaining factors. In other words, the correlation coefficient is zero. However, it is highly correlated with the JSE Actuaries All Share Index returns, the market index from which the residual market factor is derived, suggesting that it retains most of the properties of the market index.[38] The lack of correlation with the remaining factors is especially important within the context of the APT. Correlation between factors that feature in the linear factor model may result in multicollinearity that, as Blanchard (1987: 449) asserts, "is likely to prevent the data speaking loudly on some issues, even when all of the resources of economic theory have been exhausted." Williams, Grajales and Kurkiewicz (2013: 11) state that multicollinearity can lead to unstable coefficient estimates, as a result of inflated standard errors and confidence intervals.

[37] Van Rensburg's (1996: 106) approach suggests that pervasiveness is defined as having an impact on the market index, which is representative of the aggregate market. This argument has merit; the market index is assumed to represent the market, and therefore, should be devoid of unsystematic components, which have been diversified way. By this reasoning, factors identified in this manner should be representative of systematic influences.

[38] A criticism that has been levelled at this approach is that orthogonalised factors are not the original factors of interest (Wurm & Fisicaro, 2014: 42).

Burmeister and Wall (1986) demonstrate an early empirical application of the residual market factor and its efficacy. However, the residual market factor does have a theoretical basis. With a few simplifications, it can be shown that if the linear factor model omits relevant factors, and consequently, the residuals exhibit contemporaneous correlation, then the residuals of equation (2.1) can be decomposed as follows (Burmeister & McElroy, 1991; Van Rensburg, 2000: 36):

$$\varepsilon_{it} = +\sum_{j=1}^{J} b_{ij} f_{jt} + \varepsilon_{it}^{*} \tag{3.3}$$

where $\sum_{j=1}^{J} b_{ij} f_{jt}$ represents factor j or a set of factors, with the associated sensitivities, that are either unobserved or omitted from the linear factor model in equation (2.1), such that $k \neq j$, where k represents factors that are incorporated into the linear factor model. The residual term, ε_{it}^{*}, in equation (3.3) should now be uncorrelated across assets after accounting for the influence of the omitted and unobserved factors. Substituting equation (3.3) into the linear factor model in equation (2.1) leads to the following representation of the linear factor model:

$$R_{it} = E(R_i) + \sum_{k=1}^{K} b_{ik} f_{kt} + \sum_{k=1}^{J} b_{ij} f_{jt} + \varepsilon_{it}^{*} \tag{3.4}$$

where in the linear factor model underpinning the macroeconomic APT, $\sum_{k=1}^{K} b_{ik} f_{kt}$ is a set of pre-specified macroeconomic factors and associated sensitivities, and $\sum_{j=1}^{J} b_{ij} f_{jt}$ is a set of as yet unobserved and omitted factor(s) and associated sensitivities.

The relevant question now relates to the identity of the set of factor(s) *j.* One approach is to use a factor analytic augmentation, namely to apply principal component analysis or factor analysis to the residual correlation matrix of equation (2.1) to derive factor scores, and to incorporate these in place of the unspecified factor set in equation (3.4). This constitutes a factor analytic augmentation (Van Rensburg, 1997: 63). The other solution that leads to the use of a residual market factor is to use a well-diversified portfolio if there is a single unobserved factor or to replace the unobserved factors in equation (3.4) with *j* well-diversified portfolios (Burmeister & McElroy, 1991: 44). McElroy and Burmeister (1988: 33) show that this is indeed the case by referring to equation (3.3) and taking an error-components perspective. It is assumed that if the residuals, ε_{it} , in equation (3.3) are those of a well-diversified portfolio, then ε_{it}^{*} will be a degenerate factor with a magnitude of 0. Therefore, ε_{it} will now be equal to f_{jt} (assuming that b_{ij} is equal to 1) and ε_{it}^{*} is diversified away. Consequently, in equation (3.4) and if R_{it} is a well-diversified portfolio, $\sum_{j=1}^{J} b_{ij} f_{jt}$ will now represent the residuals of equation (3.4) as ε_{it}^{*} is now zero (Chang, 1991: 379). McElroy and Burmeister (1988: 33) argue that such a well-diversified portfolio will be some broad market index and therefore, a market aggregate is the basis of the residual market factor.

Further rationale for the use of the residual market factor as a proxy for omitted factors, derived from a market index, is also provided by Born and Moser (1988). Born and Moser (1988: 289) argue that the true market portfolio is an aggregation that reflects all underlying return generating factors. This argument, of course, implies that a broad market index represents the true market portfolio, or at the very least is an imperfect approximation of the true market portfolio. Nevertheless it then follows that if $R_{it} = R_{Mt}$ in equation (3.4), and R_{Mt} represents returns on a well-diversified aggregate market index, then ε_{it}^{*} is equal to zero and therefore $\sum_{j=1}^{J} b_{ij} f_{jt}$ represents the residuals of this equation, such that $\varepsilon_{it} = \sum_{j=1}^{J} b_{ij} f_{jt}$ in equation (3.3). The residuals, ε_{it} are now the residuals of a market index, so that $\varepsilon_{it} = \varepsilon_{Mt}$ and $\varepsilon_{Mt} = M\varepsilon_{t}$ in equation (3.1). Therefore, as the residual market factor represents omitted factors and reflects all underlying return generating factors and ε_{it}^{*} is now zero, the residuals of a linear factor model that incorporates the residual market factor, should be uncorrelated across assets (equation 2.2) as all remaining influences are captured in the residuals of a regression of the returns on a market index onto a set of pre-specified factors.

Wei (1988: 888) outlines some notable implications of using a residual market factor in the APT relation:

$$E(R_i) = \lambda_0 + \sum_{k=1}^{K} \lambda_k b_{ik} + \lambda_{M\varepsilon} b_{iM\varepsilon} \tag{3.5}$$

where equation (3.5) is an extension of the APT relation in equation (2.4) but incorporates the residual market factor beta, $b_{iM\varepsilon}$ for series *i*, and the associated risk premium, $\lambda_{M\varepsilon}$. Wei (1988) shows that if factors are omitted from the linear factor model, expected returns remain a function of the factor betas derived from the linear factor model and also the residual factor beta. Conversely, if the number of factors in the linear factor model approaches the true number of factors, and therefore enters the APT relation, $\lambda_{M\varepsilon}$ will decrease to a constant and will remain positive. Therefore, $\lambda_{M\varepsilon}$ is an increasing (decreasing) function of the number of omitted (included) factors; the importance of the residual market factor increases (decreases) as the number of omitted factors increases (decreases) (Wei, 1988: 888). Wei (1988) further argues that the incorporation of a residual market factor allows for an indirect test of the number of factors in the return generating process. A rejection of the hypothesis that the risk premium on the residual market beta is insignificant in the APT relation implies that the number of unobserved factors in the linear factor model is not zero, and that the number of factors in the linear factor model is insufficient or that idiosyncratic influences have not been completely diversified away in the market portfolio or both (Wei, 1988: 889).

One implication for this study is that if the residual market factor or even a second residual market factor is statistically significant in the linear factor model, then the factor set used does not adequately characterise the return generating process. A second implication is that if any other residual market factors or statistical factors derived from the residuals of a linear factor model comprising a specific factor set and a conventional residual market factor are statistically significant in addition to the factors (the residual market factor and the factor set), then the conventional residual market factor fails to account for all omitted factors. This motivates for a test of the adequacy of the residual market factor to account for omitted factors. Such a test requires that a second orthogonal residual market factor is statistically insignificant and that there are no common factors that are relegated to the residual correlation matrix. This study advocates for such a test, using a second residual market factor derived from a widely used international market index, the MSCI World Market Index. The motivation for using such an index, and this specific index, is discussed in greater detail in Chapter 4.

3.3. ALTERNATIVES TO THE RESIDUAL MARKET FACTOR

The use of the residual market factor to resolve the broader consequences of factor omission bias is common in the APT framework (Section 3.4.). However, this is not the only approach. One approach is to use techniques such as the non-linear seemingly unrelated regression (NLSUR) or non-linear three-stage least squares regression (NL3SLS) technique if a specification is assumed to be underspecified. McElroy and Burmeister *(*1988: 29-30) state that the use of NLSUR permits for the estimation of the APT in the presence of unobserved factors, mitigates efficiency losses and also mitigates the error-in-variables (EIV) problem. NL3SLS permits an approximate factor structure, which allows for correlation between the residual terms (Clare, Priestley & Thomas, 1997a: 560-561). Clare, Priestley and Thomas (1997b: 652) report that the use of NL3SLS estimators assuming an approximate factor structure bestows efficiency gains to estimated risk premia and shows that the specification of the structure of the covariance matrix is important in tests of the APT.

Van Rensburg (2002: 92), however, argues that the NLSUR approach will not confer efficiency benefits if the same set of explanatory factors is used across specifications and no across restrictions are imposed. Another approach is the use of instrumental factors and the NL3SLS approach satisfies the requirements of an instrumental estimator (Leightner & Inoue, 2012: 2; Greene, 2012: 372). Van Rensburg (2002) is critical of the challenge associated with selecting appropriate instrumental variables and the loss in efficiency associated with instrumental variable techniques (also see Gujarati, 2004: 678).[39] Leightner and Inoue (2012: 22)[40] propose the use of yet another econometric technique, the best projection reiterative truncated projected least squares (BP-RTPLS) approach. This approach uses the vertical position of observations to account for omitted factors that interact with included factors. It is shown that the technique produces estimates that are unbiased and does away with the need for proxy factors or instruments. However, this technique does not quantify the impact of unknown factors because it ignores information. Van Rensburg (2002) argues that the use of a residual market factor avoids the abovementioned problems associated with simultaneous equation techniques and suggests that use of a residual market factor (or factors) is an appropriate solution for all linear factor models that employ pre-specified factors. It also permits the sources of unobserved variation to be explicitly reflected in the linear factor model and permits their quantification. Therefore,

[39] Gujarati (2004: 678) defines instrumental variables as factors that are correlated with a given explanatory factor but uncorrelated with the residuals. It is easy to imagine the challenges involved in trying to find such a factor. This is especially true if the structure of the return generating process is not known.
[40] See Leightner and Inoue (2012) for a discussion of a number of related techniques.

the use of a residual market factor (or factors) also potentially reveals important information about the return generating process (Section 3.4.).[41]

The contribution of the residual market factor to the APT is two-fold. The residual market factor acts as a proxy for unobserved and omitted factors. It permits the explicit representation of these factors in the linear factor model. As suggested in Wei (1988: 889), the risk premia on the residual market factor beta is indicative of the magnitude of the underspecification as it is a positive function of omitted factors. This constitutes an indirect test of the number of factors in the return generating process. It also permits a breakdown of the pervasive influences that impact stock returns without affecting the estimated coefficients on the macroeconomic factors that are used to proxy for the underlying pervasive influences in stock returns while accounting for omitted factors.

3.4. THE RESIDUAL MARKET FACTOR IN APT LITERATURE

Berry *et al.* (1988)[42] employ the residual market factor in a study of the exposures of returns on US economic and industrial sectors to four macroeconomic factors representative of pervasive influences. The fifth factor is the residual market factor, derived from returns on the S&P 500 Index. Berry *et al*. (1988: 31) motivate for the inclusion of the residual market factor by postulating that the "worry over possible missing factors is substantially resolved by using a residual market factor" and that the residual market factor embodies all factors in the same manner that the market index embodies all factors in the market model or the CAPM. This aptly articulates the perceived role of the residual market factor in the literature and in the broader application of the APT. Moreover, it is also shown in the derivation of the residual market factor, that returns on the S&P 500 Index are significantly related to the remaining factors that feature in the proposed specification, namely the default spread, the term structure, inflation and unexpected changes in the growth rate of profits. Chan *et al.* (1985: 452) suggest that it is sufficient to show that information is reflected in a given market index by showing that there is a correlation between a market index and macroeconomic

[41] A caveat here is that Van Rensburg's (2002) study concerns the industrial and resource dichotomy that underlies the return generating process underlying the JSE and the solution suggested relates to linear factor models for the JSE. Of course, such a dichotomy may exist in other markets and therefore, a single residual market factor derived from the domestic market index may be insufficient to account for omitted factors.

[42] Berry *et al.'s* (1988) study is also discussed in Section 2.4. The study is a seminal example of how the APT can be applied to study the return generating process of stock returns. The present discussion of this study in this section emphasises the role residual market factor.

factors. Berry *et al.* (1988) show that macroeconomic information is reflected in the market index and confirm the role of the market index as a proxy for macroeconomic information.

McElroy and Burmeister (1988: 29) apply the iterated NLSUR methodology to jointly estimate factor coefficients and the associated risk premia (this is in contrast to the Fama-Macbeth two step procedure). As in Berry *et al.* (1988), returns on the S&P 500 Index are regressed onto four macroeconomic factors, namely the default spread, the term structure, unexpected deflation and real final sales. The results indicate that the four macroeconomic factors and the residual market factor are important in explaining returns (215 of 350 coefficients are statistically significant) and all macroeconomic factors are priced. That the residual market factor is not priced indicates that its role is confined to explaining time series variation in this study. The authors go on to investigate whether the four factors contribute additional information that is not already reflected in the returns on the S&P 500 Index. Restricted (S&P 500 Index only) and unrestricted (all factors) models are estimated and the resultant residual covariance matrices are compared. Results indicate that the macroeconomic factors are essential for explaining realised returns; the contemporaneous covariance matrices derived from these models differ significantly. Finally, McElroy and Burmeister (1988: 41) suggest a role for the residual market factor that goes beyond that of measuring omitted factors. The authors argue that although there may be other sets of macroeconomic factors that perform better than the factor set used and that some factors may be missing from the specification, these influences are captured by the factor set used and the residual market factor. Therefore, the residual market factor should complement any macroeconomic factor set to yield an adequate description of the return generating process.

Chang (1991: 380) argues that employing a market index in a multifactor linear model can lead to two possible outcomes. The first is a severe multicollinearity problem which will render coefficients erroneously insignificant. The second is a significant result for the market index, indicating that there are factors that are omitted from the linear factor model. It is further postulated that a residual market factor should be considered in the linear factor model and that this residual market factor will reflect information that is required for a correct specification of the underlying factor structure. Chang (1991) further investigates the role of the residual market factor from a pricing perspective by comparing the results of models that include a market index and, alternatively, a residual market factor and only macroeconomic factors. Results indicate that in the macroeconomic factor model, the risk premia associated with the six macroeconomic factors[43] considered are statistically significant across subperiods, in the absence of a market beta. When a market beta is included in the underlying multifactor model, pricing significance shifts away from some of the macroeconomic factors to the market factor and a number risk premia associated with the macroeconomic factors now have the wrong signs. This is attributed to suspected multicollinearity whereas the significance of the market beta is associated with underspecification of the underlying linear factor model.[44] When the residual market beta is used, the factor risk premia regain their significance and the expected signs, while the residual market beta is statistically significant. Chang (1991) cites this as evidence that market residuals reflect missing factor information and argues that the residual market factor should enter the APT framework to reflect the influence of unobserved and missing factors. These findings provide support for the derived propositions of Wei (1988) and demonstrate the residual market factor's theoretical (that of incorporating omitted information) and econometric (that of mitigating potential multicollinearity associated with the inclusion of a market index) contributions.

Koutoulas and Kryzanowski (1994: 332-333) investigate the integration (or lack thereof) of the Canadian and North American equity markets by arguing for an APT specification that incorporates both domestic and international factors. A version of the APT that includes three international factors and three domestic factors is estimated using returns on 50 equally-weighted size-ranked portfolios of Canadian stocks.[45] Both domestic and international factors are priced, providing support for partial integration. This specification is then augmented with a residual market factor, which Koutoulas and Kryzanowski (1994: 344) argue captures the influence of any omitted domestic and international factors. The risk premium on the residual market factor is statistically significant. Consequently, the authors suggest that there may be omitted domestic and/or international risk factors that are (also) priced although it is not known whether the residual market factor reflects domestic or

[43] The six factors are industrial production, changes in the yield structure, unanticipated inflation, unemployment, the default risk premium and the exchange rate.

[44] This demonstrates how even low levels of correlation between macroeconomic factors have the potential to result in multicollinearity and lead to misleading inferences (Williams *et al.*, 2013). Chang (1991: 383) reports that the correlation between the market index and the yield structure is slightly higher than 0.3 yet the yield structure loses its pricing significance during some subperiods. This suggests that although correlation between factors is not particularly high, multicollinearity may still be problematic.

[45] These factors are the pure domestic components of industrial production, an index of leading domestic indicators and the term structure and the pure international components of rates on Euro-Dollar deposits, differentials between Canadian and US leading indices and the differential between Canadian and US industrial production (Koutoulas & Kryzanowksi, 1994: 344).

international influences. The study is an application of the residual market factor in asset pricing but raises an important question. This is the question of whether the residual market factor can account for both domestic and international influences. The study also suggests that if the residual market factor is a proxy for omitted factors, then a global market index that is orthogonal to the conventional residual market factor should be irrelevant. This provides the basis for a two residual market factor approach and a test of the adequacy of the conventional residual market factor. If a second residual market factor derived from a global market index is statistically significant, then the conventional residual market factor does not reflect all global macroeconomic influences.

Kryzanowski, Lalancette and To (1994: 155-156) investigate whether the inclusion of the residual market factor resolves mispricing in the APT. The authors propose that if the residual market factor is zero, then the factor structure reflects the true number of factors and the market portfolio is perfectly diversified. If this is not the case, then the APT will reflect mispricing unless the residual market factor is incorporated. In the first test, Kryzanowksi *et al.* (1994) regress returns on individual stocks onto one, six and eight statistically derived factor structures. The residual series are then regressed onto a residual market factor constructed from returns on the CRSP value-weighted index. Results indicate that the residual market factor is statistically significant for over a third of stocks for a one-factor structure, but for under a fifth of stocks when a six factor structure is assumed. Additionally, after adjusting covariance matrices for nonsynchronous trading, the residual market factor is significant for almost 40% of stocks when a one factor structure is assumed. This declines to under a third when an eight factor structure is assumed.[46] Kryzanowski *et al.* (1994) note that the number of stocks for which the residual market factor is significant is higher when fewer factors are used, suggesting that additional factors account for some omitted factors, but not all. Cross-sectional tests yield somewhat different results. Six and eight factor structures appear to ensure an exact factor structure and a single statistical factor APT produces statistically insignificant pricing errors indicating that the residual market factor is irrelevant in determining cross-sectional returns. These findings are somewhat similar to those in McElroy and Burmeister (1988). The results imply that although the residual market factor may not be necessary to resolve underspecification in the APT relation, it plays a role

[46] Kryzanowski *et al.* (1994: 170; 172) adjust the covariance matrices for nonsynchronous trading. When adjustments are made, the residual market factor remains significant for both one and eight-factor structures, indicating that an additional seven factors are unable to account for omitted influences, which are then reflected by the residual market factor.

in the linear factor model. It is possible that results will differ if factors are of a macroeconomic nature and not of a statistical nature.

Van Rensburg (1995: 55-56) investigates the relationship between macroeconomic factors and returns for the South African gold (mining), financial and industrial sectors and a combination of the mining-financial sectors, within the APT framework. Four factors, namely unanticipated changes in the gold price, returns on the Dow Jones Industrial Index, inflation expectations and the term structure of interest rates are considered. The residual market factor is derived by regressing returns on the JSE All Share Index onto these factors. Van Rensburg (1995) emphasises the role of the residual market factor by stating that it captures the variation in returns that is not explained by the remaining factors and serves a "catch-all" purpose. The gold and mining-financial sector indices are more sensitive to the residual market factor than the financial and industrial sector indices (coefficients of 0.947 and 0.972 vs 0.527 and 0.630 respectively). According to the author, this implies that factors that have explanatory power for these sectors have been omitted and the effect of these factors is reflected in the residual market factor to a greater extent. While Wei (1988) postulates that the residual market factor premium in the APT relation is an increasing function of omitted factors, Van Rensburg (1995) suggests that this interpretation is applicable to the estimated sensitivity to the residual market factor in the linear factor model.

In a subsequent study, Van Rensburg (1997: 61-62) extracts two factors[47] from the residuals of a linear factor model relating returns on industrial, financial, mining and mining-financial stocks to returns on the Dow Jones Industrial Index, unanticipated changes in inflation expectations and unanticipated changes in the term structure. It is argued that the subsequent inclusion of the extracted factors in the linear factor model constitutes a factor analytic augmentation that ensures that residuals are uncorrelated across stocks. Stepwise regression is then applied to identify priced factors in addition to pre-specified macroeconomic factors using the iterated NLSUR approach. The three factors that compete for entry into the APT relation are the two statistical factors and the residual market factor, derived by regressing returns on the JSE All Share Index onto the three pre-specified factors. In the first round, the three pre-specified factors are priced, together with the first statistical factor, which Van Rensburg (1997) suggests is representative of an industry

[47] Van Rensburg (1997: 65) states that these two factors represent broad industrial and gold/mining influences and the existence of these factors provides support for a gold-industrial dichotomy underlying the South African stock market.

effect. After excluding the first derived factor, the second statistical factor enters the APT model, but is not priced. In the final round, the residual market factor is included and is priced after both statistical factors are excluded. This indicates that the first statistical factor, aside from the residual market factor, is important for pricing and reflects omitted influences. A finding that the residual market factor competes with this factor suggests that both factors reflect influences that would otherwise be reflected in the residuals.[48] It also suggests that the residual market factor may not reflect all relevant information for pricing and therefore may not be an adequate proxy for omitted factors. Van Rensburg (1997) argues for the inclusion of this statistical factor in the linear factor model. Although these findings provide evidence that the residual market factor reflects influences in the residuals of the linear factor model, they also suggest that the residual market factor may not be an optimal proxy for omitted influences.

Clare and Priestley (1998: 104) study pricing in the Malaysian stock market and argue that the removal of capital barriers allows for the possibility of domestic and international factors to feature in pricing relationships. Aside from factors representative of domestic risk, namely unexpected changes in the risk free rate, the term structure, industrial production, unexpected inflation and expected inflation, the APT model is extended to include returns on the Kuala Lumpur Composite Index to capture omitted domestic factors. Additionally, to measure international influences, a residual market factor derived from the MSCI World Market Index is also included.[49] Results indicate that both factors are priced, suggesting that two proxies may be required to reflect omitted factors, with one of these being international in nature. Parameters are estimated jointly using the NLSUR approach and Clare and Priestley (1998) compare the $\bar{R}^2$s of a domestic version of the APT (which excludes the second residual market factor) and that of the international APT. The average $\bar{R}^2$ increases from 45.894% to 48.412% indicating that the inclusion of the international residual market factor improves model specification, albeit marginally. It also suggests that specifications motivated by the APT may benefit from two residual market factors and that a single residual market factor may not be sufficient to account for all influences.

[48] In a second round of tests, the pre-specified factors are included together with the first statistical factor and a respecified residual market factor. The respecification takes the form of a regression of market returns onto the three pre-specified factors and the statistical factor. Both the statistical factor and the residual market factor are not priced.

[49] The derivation of the international residual market factor is somewhat different in this study from the conventional derivation. The Kuala Lumpur Composite Index is orthogonalised by regressing the returns on this index onto returns on the MSWCM index (Clare & Priestley, 1998: 111).

Van Rensburg (2000: 32) motivates for the use of two residual market factors to resolve potential model underspecification arising from a dichotomy in the return generating process underlying South African mining and industrial stocks (also see Van Rensburg, 1997; 1998; Van Rensburg & Slaney, 1997). The author argues that contemporaneous pairwise residual correlation is driven by underspecification and that the solution is to derive statistical factors from the residuals and to incorporate these into the linear factor model. Moreover, Van Rensburg (2000) further argues that information relegated to the residuals may be utilised by employing two industrial sector indices, the JSE Industrial and Gold Indices, as proxies for these unspecified factors. Subsequently, returns on these indices are used to derive two residual market factors. This is in line with Burmeister and McElroy's (1991) proposition that well-diversified portfolios may be used to replace unobserved factors that are relegated to the residuals (Section 3.2.). Both residual market factors are found to contribute significantly to the explanation of the return generating process underlying returns on the JSE All Share Index. The $\bar{R}^2$ increases from 0.29 to 0.91 when returns are regressed onto innovations in All Gold Index earnings, returns on the Dow Jones Industrial Index, the 10 year government bond yield, the level of gold and foreign exchange reserves and the Rand gold price *and* these two factors. In the cross-section, both residual market factors are priced. Van Rensburg (2000: 41) concludes by stating that the use of two residual market factors avoids the misspecification error that is introduced by a mining and industrial dichotomy in the linear factor model that describes the return generating process of South African stock returns. Similarly to Clare and Priestley (1998), these results indicate that a single residual market factor may not be sufficient to account for omitted factors in the linear factor model and the APT relation. However, in contrast to Clare and Priestley (1998), it appears that this is driven by a specific characteristic, an industrial dichotomy, of the return generating process underpinning the South African stock market as opposed to relevant international influences that may or may not be reflected in a residual market factor constructed from an aggregate domestic market index.

In their study of pricing in international markets, Brown *et al.*(2009: 296, 298) note that equity market correlations have increased around the world. In doing so, residual market factors are applied to model national market indices. In estimating an international linear factor model for 21 national markets, two residual market factors derived from a global equity index, the MSCI World Market Index, and a bond market index, the Citigroup World

Government Bond Index, are used in addition to a number of global factors.[50] Both residual market factors feature prominently (are statistically significant) in the return generating process for the markets considered in the sample. The use of two residual market factors derived from purely international indices is conceptually similar to that in Clare and Priestley (1998) and Van Rensburg (2000). Notably, the second residual market factor is not derived from an equity-based series but from a bond index. This application again suggests that more than one residual factor may be required to capture the influence of omitted factors.

The literature shows that a residual market factor is widely considered to be a proxy for omitted factors. It is proposed, notably by Berry *et al.* (1988) and Kryzanowksi *et al.* (1994), that the inclusion of this factor resolves underspecification of the linear factor model and mispricing in the APT. Wei (1988) demonstrates the theoretical contribution of the residual market factor to the APT relation in the form of an increasing risk premium associated with omitted factors. Chang (1991) demonstrates the econometric contribution of the residual market factor in the form of a formulation that does not cause multicollinearity. Van Rensburg (1995) argues that the sensitivity of the residual market factor quantifies the influence of omitted factors in the linear factor model and Van Rensburg (1997) provides evidence that the residual market factor proxies for omitted factors, although it may not be a perfect proxy for all omitted influences. Clare and Priestley (1998) depart from the use of a single residual market factor. Van Rensburg (2000) and Brown *et al.* (2009) also depart from the use of a single residual market factor, albeit for different reasons. The use of two residual market factors opens a further avenue of consideration, that of whether a single residual market factor is sufficient. This aspect is explored in this study. It is argued that a second residual market factor should be redundant if the conventional residual market factor is an adequate proxy for omitted factors. Consideration is also given to whether a second residual market factor can adequately resolve any remaining factor omission bias.

3.5. INFORMATIONAL CONTENT

Chan *et al.* (1985: 452) state that if markets are efficient, information is quickly reflected in the market aggregate and correlation between the market aggregate and measures of macroeconomic activity is evidence of this (also see Fama, 1965; 1970; 1981; 1995). This

[50] These are the crude oil price, returns on small US stocks relative to the MSCI World Market Index returns, Euro one-month money market returns, changes in the Yen-Dollar exchange rates and the Fama-French HML factors (Brown *et al.,* 2009: 299). Also see Aretz, Bartram and Pope (2010), who suggest that Fama and French (1993) factors are proxies for macroeconomic fundamentals.

reasoning stems from Fama's (1970) efficient market hypothesis that markets are extremely efficient in reflecting information and therefore any new information is rapidly reflected in stock prices (also see Malkiel, 2003: 59). This reasoning extends to the residual market factor, which is derived from the market index and is therefore also an indicator of changes in economic conditions and sentiment (Spronk & Hallerbach, 1997: 123; Osamwonyi, & Evbayiro-Osagie, 2012: 55).

At a very abstract level, Kwon and Yang (2008: 2854) show, by applying the (physics) concept of transfer entropy, that there is a bi-directional flow of information between stock market indices (S&P 500 Index and Dow Jones) and individual stocks implying that market indices are both aggregators and transmitters of information. Cutler, Poterba and Summers (1989: 5) investigate the impact of economic news on stock returns using a value-weighted NYSE index, which they refer to as a "news proxy." Results indicate that the index reflects both current and prior economic developments, although, current economic developments have a greater impact.[51] While it is shown that macroeconomic news explains a significant amount of variation in returns, it does not explain all variation. This indicates that, while considered a proxy for macroeconomic news, the market index may also account for other developments. Cutler *et al.* (1989) show that this is indeed the case by relating returns on the S&P Index to non-economic events and international conflicts (e.g. Pearl Harbour in 1941 and Kennedy's assassination in 1963) and note large changes in the level of this index.

McQueen and Roley (1993: 694) show that the response of stock prices to macroeconomic information differs during different stages of the business cycle. Out of a set of eight factors, returns on the S&P 500 Index are significantly related to changes in the producer price index (PPI) and money supply ($\bar{R}^2$ = 0.02) over the entire sample period between September 1977 and May 1988. The S&P 500 Index performs better in reflecting macroeconomic information during favourable economic conditions. The S&P 500 Index reflects announcements relating to industrial production, unemployment, the merchandise trade deficit, the producer price index and the money supply. However, when economic activity is in a medium state (as defined by McQueen & Roley, 1993), the S&P 500 Index responds to announcements about inflation and the producer price index and does not respond to any announcements during states of low economic activity. While this is evidence that a market

[51] Cutler *et al.* (1989: 6-7) show this by using contemporaneous, lagged and led macroeconomic factors in an unrestricted VAR model. The addition of contemporaneous and lagged factors improves the $\bar{R}^2$ substantially. The macroeconomic factors considered are industrial production, real money, interest rates and inflation.

index reflects macroeconomic news, it indicates that a market index is not an adequate proxy for economic news at all times.

Flannery and Protopapadakis (2002: 773-774) study the impact of an extensive set of macroeconomic announcements[52] on the returns on a value-weighted NYSE-AMEX-NASDAQ index. Significant mean returns are associated with five types of announcements relating to the balance of trade, construction spending, unemployment, personal consumption and the PPI. Additionally, announcements for the balance of trade, inflation, employment, housing starts, the money supply and the PPI are found to be associated with increased trading volumes, confirming that markets react to news announcements (Flannery & Protopapadakis, 2002: 762-763). This again demonstrates the role of market indices as an aggregator of information. Baker and Wurgler (2007: 138) investigate whether investor sentiment is reflected in market aggregates by constructing a sentiment index consisting of six factors[53] representative of sentiment and then establishing the relationships between this index and returns on funds. A sizable correlation between sentiment and returns is reported; the authors find significant contemporaneous correlation of 0.43 between returns on an equally-weighted market index and changes in the sentiment index suggesting that sentiment is reflected in this index. This finding goes beyond standard finance theory according to which stock prices are determined by rational investors on the basis of expected cash flows (Baker & Wurgler, 2007: 129). Birz and Lott (2011: 2792) study the impact of macroeconomic news by investigating the effect of newspaper stories relating to US Gross Domestic Product (GDP) growth, unemployment, retail sales and durable goods on stock prices. Returns on the S&P 500 Index are found to reflect newspaper headlines relating to the GDP growth rate and unemployment rate but not retail sales and sales of durable goods. Birz and Lott (2011) attribute this to the relative low importance of these factors in informing investors' expectations of future economic conditions. Although this again supports the proposition that a market aggregate reflects macroeconomic news, it also indicates that not all information is reflected.

The literature proposes that market aggregates are indicators of changes in economic conditions and aggregate information about the state of the economy. As suggested by

[52] Flannery and Protopapadakis (2002: 774) state that the set of macroeconomic announcements is the most extensive dataset evaluated over the 1980 to 1996 period.
[53] These are trading volume as measured by NYSE turnover, the dividend premium, the closed-end fund discount, the number and first-day returns on IPOs and equity share in new issues (Baker & Wrugler, 2007: 138).

Baker and Wurgler (2007), sentiment is also reflected in market aggregates. These findings are in line with the proposition that the market index, and specifically the residual market factor, will reflect omitted factors. According to McQueen and Roley (1993), the effect of macroeconomic news on the market aggregate varies according to the economic state. This suggests that a market index may not always reflect macroeconomic information. Similarly, Birz and Lott (2011) suggest that not all information is reflected in market aggregates. This constitutes an early as yet unexplored insinuation, together with studies (discussed in Section 3.4.) that incorporate two residual market factors, namely that a single and domestic market index and the derived residual market factor may not be adequate proxies for all omitted factors.

3.6. CHAPTER SUMMARY AND CONCLUSION

The residual market factor, introduced by Burmeister and Wall (1986) and expounded theoretically by Burmeister and McElroy (1991), is a construct that is uncorrelated with the factors included in a linear factor model specification, which is applied across a number of assets. It is postulated that the residual market factor will be represented by a well-diversified portfolio, such as an aggregate market index (Section 3.2.). Although there are other approaches to resolving underspecification in the linear factor model, the use of a residual market factor potentially conveys important information about the return generating process and is easily implementable (Section 3.3.).

Literature has readily adopted the residual market factor as a proxy for omitted factors and it is hypothesised that the inclusion of the residual market factor will resolve possible underspecification. The literature suggests that the residual market factor makes an important theoretical and empirical contribution to the estimation and interpretation of the linear factor model and the APT relation (Section 3.4.). Section 3.5. demonstrates that market aggregates reflect macroeconomic news and investor sentiment. The literature discussed in Section 3.5. indicates that market aggregates are a proxy for macroeconomic information and also sentiment. Therefore, a residual market factor should fulfil the role of a proxy for the multitude of omitted factors, both observed and unobserved, that impact returns.

Two themes emerge from this chapter. The first is that the approach of using a residual market factor to reflect omitted and unobserved factors, is widely applied in the literature. This study interrogates the proposition that the inclusion of the residual market factor will resolve underspecification that potentially arises in macroeconomic linear factor models. The second theme is that two residual market factors may need to be considered. As suggested in Section 3.5., this is because either a single residual market factor may not adequately capture dichotomies in the return generating process or because there are international influences or there are other factors that are not reflected by a single residual market factor. Regardless of the reason, this study does not seek to argue for the use of a second residual market factor as a standard approach. Rather, it aims to determine whether a single residual market factor is an adequate proxy for omitted factors by establishing whether a second orthogonal residual market factor significantly features in the linear factor model. If such a factor features in the linear factor model, then a single residual market factor derived from a domestic market aggregate is not an adequate proxy for omitted factors.

Chapter 4 explores the role of international influences in returns and motivates for the inclusion of a second residual market factor as a test factor. Such a factor, as it is shown in the following chapter, can be derived from returns on the MSCI World Market Index, an index that is widely used to proxy for international influences.

CHAPTER 4

INTERNATIONAL INFLUENCES AND THE ARBITRAGE PRICING THEORY

4.1. INTRODUCTION

This chapter investigates the role of international influences in stock returns and seeks to understand the role of international influences in the APT. International influences in stock returns emerge as a theme in Chapter 2 through the extension of the APT to the International APT (Section 2.3.2.). International influences also feature in Chapter 3 and are captured by an international equity index used alongside a domestic market index or residual market factor (Section 3.4.).

There is little doubt that linkages exist between national stock markets as exemplified by the crash of October 1987. The removal of legislative barriers and increased capital mobility have given rise to the importance of international factors in the risk-return relationship (Van Rensburg, 1995: 49; Clare & Priestley, 1998: 104). Therefore, it follows that international influences should be considered in the linear factor model. Moreover, the importance of international influences offers the opportunity to assess the adequacy of a single residual market factor. If a single residual market factor derived from the domestic market aggregate is a sufficient proxy for omitted and unobserved influences, then a second residual market factor should not feature in the linear factor model.

A global market index, acting as a proxy for international influences, presents a viable candidate for the derivation of a second residual market factor. This chapter argues for the consideration of a specific market index, the MSCI World Market Index, in the role of a second residual market factor. As demonstrated in this chapter, the MSCI World Market Index is widely used to proxy for global factors in returns. Therefore, this factor is used to derive a second residual market factor which is treated as a test factor to determine whether the first residual market factor adequately captures international influences. If this is not the case, the inclusion of this factor permits a consideration of whether a two residual market factor approach resolves underspecification.

Section 4.2. motivates for the consideration of international influences in returns and rationalises the role of international influences in the context of news and information spillovers. As with the residual market factor, this indicates that a proxy for international influences should reflect news about the global macroeconomic state. Section 4.3. outlines how the APT approaches global influences and international risk. Following this, the informational content of an international equity index, the MSCI World Market Index, is outlined in Section 4.4. analogously to that of the conventional residual market factor (Section 3.5.). Section 4.5. summarises and concludes this chapter.

4.2. INTERNATIONAL INFLUENCES IN RETURNS

4.2.1. Interdependence

Immediate evidence of the impact of common global influences in stock returns is stock market co-movement. Eun and Shim (1989: 242) state that international stock market movements indicate that there is interdependence between national markets and that developments in these markets in the form of news events impact individual national markets. In preliminary analysis, the contemporaneous correlations and interdependence across nine markets are considered.[54] Eun and Shim (1989) report that intra-regional correlations (for example, US/Canada, Germany/Switzerland and Hong Kong/Japan) are higher relative to inter-regional correlations (for example, Canada/Japan and France/Hong Kong) and while all markets are correlated, correlation with the US declines the further a market is away geographically. The observed correlation patterns are attributed to differing levels of economic integration between countries. Importantly, results from a Vector Autoregression (VAR) analysis indicate that no market is exogenous. The average percentage of variance of a given market that is attributable to innovations in foreign stock markets is 25.93% and ranges between 11.02% for the US and 52.02% for Canada. The US market is the most influential market, explaining on average 16.78% of variance in foreign markets. The authors conclude that these findings point towards substantial interdependence between national stock markets. This implies that news emerging from national markets impacts other external markets.

Bradfield (1990: 3) finds that the US stock market is the most internationally influential stock market for returns on the JSE. Correlation coefficients show that the JSE, Tokyo Stock Exchange (TSE) and the London Stock Exchange (LSE) are strongly correlated with the NYSE and that the level of correlation (of 0.268) between returns on the JSE and NYSE is higher than that of the JSE and the TSE and LSE. Bradfield (1990) regresses returns on 30 South African stocks onto Dollar returns on the DJIA and finds that almost two-thirds of stocks are related to movements of the NYSE. Stocks that are most sensitive to movements

[54] The nine markets considered in the study are Australia, Canada, France, Germany, Hong Kong, Switzerland, the UK and the US.

on the NYSE (as measured by the DJIA) are those that derive a large portion of their income from external markets. The author concludes that the empirical evidence presented suggests that a large proportion of South African stocks is influenced by movements on the NYSE. Similarly to the findings of Eun and Shim (1989), these findings show that not only is there co-movement at market aggregate level, but that interdependence is reflected at the individual asset level. Similarly, the US stock market is also the most influential market for the South African stock market as it is for other markets.

Arshanapalli and Doukas (1993: 193, 195) postulate that the 1987[55] behaviour of national markets was more influenced by international events than ever. The authors investigate the linkages between the French, German, Japanese, UK and US stock markets prior to and post the October 1987 crash. Co-integration tests, indicative of long-run linkages, suggest that while linkages between markets prior to October 1987 are weak, they have increased substantially following the post-crash period. While none of the markets in the sample are co-integrated with the US market prior to October 1987, three out of the four markets considered (France, Japan and the UK) are co-integrated after October 1987. However, Arshanapalli and Doukas (1993) do not find evidence of co-integration of the Japanese stock market with the French, German and UK stock markets before to and after 1987 suggesting that the common market to which other markets are linked is the US stock market. Results of the error-correction model (which indicates short-run linkages and long-run adjustments) for the post 1987 period indicate that three stock markets (the French, German and UK stock markets) respond to movements of the US stock market and that the response of these markets is efficient. Arshanapalli and Doukas (1993) conclude that stock market interdependence has increased significantly post October 1987 and that the markets in the sample are influenced by the US stock market index. These findings indicate that interdependence has grown over time and that markets are interrelated through a common market, namely the US stock market, analogous to a common factor.

Similarly to Arshanapalli and Doukas (1993), Hassan and Naka (1996: 387-388) investigate short-term and long-term relationships between the US, Japanese, UK and German stock markets.[56] The authors propose that growing economic interdependence and policy co-

[55] The year of the "Black Monday" crash, which took place on 19 October 1987.
[56] Hassan and Naka (1996) use the error correction model (ECM), which permits the modelling of both short-run and long-run relationships. See Hassan and Naka (1996: 392) for an outline of the methodology.

ordination among countries are responsible for linkages in stock prices over the long-run and common co-movements in national income and expectations contribute to these linkages. Preliminary analysis shows that the respective market indices[57] are correlated at levels and in differences, which is evidence of similar national economic activities, according to Hassan and Naka (1996). The authors find evidence of co-integration amongst the stock market indices considered and attribute this to comovement driven by information sharing and the accessibility of domestic and foreign investors to these markets. The number of cointegrating relationships increases from the pre-crash period (US-Japan-UK stock markets) to the post-crash period (US-Japan-UK, US-UK-Germany stock markets) suggesting that market integration has increased following the 1987 crash. This finding is similar to that of Arshanapalli and Doukas (1993). In the short-run, Hassan and Naka (1999) report that the US stock market leads the Japanese and UK stock markets in the pre-crash period and in the post-crash period and that there are significant feedback relationships between the US, Japanese and UK stock markets. The authors conclude by stating that these results provide support for growing interdependence between these markets in the short-run and in the long-run.

Masih and Masih (1999: 254, 264) investigate linkages between southeast Asian markets and a number of other developed markets. Causality tests indicate that the US stock market is the most influential market and the US stock market is both a receptor of exogenous shocks and transmitter of information to Asian, notably the Singaporean, Hong Kongese and Malaysian stock markets and major (non-Asian) markets, namely the UK and German markets. There are also regional interdependencies between Asian markets and Masih and Masih (1999) report that both the US and UK markets are important in influencing Asian markets. The authors conclude that the findings of the study provide evidence of significant short-term and long-term interdependencies between established markets and emerging Asian markets, interdependencies that are both inter- and intra-regional. Similarly to Hassan and Naka (1996), multidirectional interdependencies point towards complex relationships between markets.

[57] The S&P 500, the Nikkei 225, the FTSE 100 and the Germany-Commerzbank Index

Collins and Biekpe (2003:182-183) investigate interdependence and contagion[58] in African markets, attributing higher responsiveness to shocks to greater financial integration. The authors investigate the effect of the Hong Kong market crash in October 1997 owing to its widespread global impact. An analysis of correlations indicates that there are strong relationships between some southern African markets as well as other African markets; [59] examples are the stock markets of South Africa and Botswana and South Africa and Egypt. The most notable finding is that two African markets, the South African and Egyptian markets, exhibit evidence of contagion following the Hong Kong market crash. Collins and Biekpe (2003) attribute evidence of the Hong Kong market crash on these markets to South Africa's and Egypt's roles as significant players in emerging markets and participation in the International Finance Corporation Investibles (IFCI) index. The findings of this study, together with those of Masih and Masih (1999), suggest that market interdependence is not only a characteristic of North American and European stock markets and also the Japanese stock market, but extends to certain African markets. This justifies a general consideration of international influences in linear factor models, regardless of the market under consideration.

The studies outlined above indicate that markets are interdependent and influenced by events in other markets. Arshanapalli and Doukas (1993: 193), Hassan and Naka (1996: 390), Masih and Masih (1999: 272) and Collins and Biekpe (2003: 182) cite the relaxations of capital controls, improvements in technology, lower costs of information flows and financial transactions, increases in cross-border operations by multinationals, deregulation, liberalisation, macroeconomic policy co-ordination, greater economic similarities and global integration as some reasons for the observed interdependencies. Hassan and Naka (1996: 390) and Masih and Masih (1999: 264) suggest that these developments have resulted in markets being responsive to international information. Therefore, while it is shown that markets are interrelated, information is ultimately responsible for market co-movements and this information is reflected in national markets because of market interdependence. The nature of this information is important; if it is macroeconomic in nature, it impacts national

[58] Collins and Biekpe (2003: 184) define contagion as the increase in market correlations following turmoil. If there is no change in correlations, then this implies that there is interdependence but not contagion between markets. In both cases, this suggests that there are linkages between markets.
[59] The markets considered are those of Botswana, Egypt, Ghana, Kenya, Mauritius, Morocco, Namibia, Nigeria, South Africa and Zimbabwe.

markets and will be reflected by an international market aggregate. The next section demonstrates that it is information that moves international markets.

4.2.2. The Role Of Macroeconomic Information And Information Spillovers

Kim (2003: 612) argues that international macroeconomic announcements provide information relating to domestic economic conditions and this is reflected in market movements. The author investigates the effects of US and Japanese macroeconomic announcements[60] on Asian markets and the Australian market. Macroeconomic announcements emanating from the US are shown to have an (overall positive) impact on returns on all markets in the sample, namely the Australian, Japanese, Hong Kongese and Singaporean stock markets. According to Kim (2003), this suggests that news emanating from the US improves market sentiment in Asia-Pacific and leads to upward revisions of earnings. Furthermore, it is also shown that there are volatility[61] spillover effects. Similarly, the US, Australian, Hong Kongese and Singaporean markets are responsive to Japanese macroeconomic news, although to a lesser extent in terms of return behaviour and volatility. Kim (2003) concludes that this evidence suggests that market participants in Asian-Pacific stock markets follow market developments in the US and Japan and that these markets incorporate the disaggregated informational flows that originate from these markets.

Wang and Firth (2004: 245, 248, 252) investigate information transmission between Chinese markets[62] and developed markets. Evidence from analysis based upon the conditional mean and variance specifications suggests that there are contemporaneous information spillovers from the developed markets, namely the Japanese, UK and the US markets,[63] to Chinese markets. These spillovers are reflected in returns and volatility. The Japanese and US markets are information leaders and the UK market is also a leader for certain markets.[64] However, volatility spillovers are bi-directional suggesting that there are also information flows between markets that are reflected in the second moment of the return distribution.

[60] The US macroeconomic information is for the balance of trade, real GDP growth rate, retail sales growth, unemployment rate, producer price inflation and consumer price inflation. The Japanese macroeconomic information is for the balance of trade, current account balance, unemployment rate, money supply, wholesale price inflation and consumer price inflation.
[61] The volatility of returns can be interpreted as a proxy for the impact of information (Armitage & Brzeszczyński, 2011: 1533)
[62] The Greater China markets are the Shanghai Stock Exchange, the Shenzhen Stock Exchange, the Hong Kong Stock Exchange and the Taipei Stock Exchange.
[63] As measured by the Nikkei 225, FT100 and S&P 500 respectively.
[64] This depends upon which Chinese market is considered. For example, the Shanghai and Shenzhen Stock Exchanges are mostly influenced by the Japanese stock market and the Hong Kong Stock Exchange and the Taipei Stock Exchange are most influenced by the US and UK markets (Wang & Firth, 2004: 245).

Wang and Firth (2004) also report that while the information spillovers are unidirectional prior to the 1997 Asian crisis, they are bi-directional following the crisis, with the US and UK markets reflecting information from the Taiwanese market and the UK market responding to the Hong Kong and Japanese markets. This is attributed to investors in other markets paying closer attention to information originating from Asian markets following the Asian financial crisis. Wang and Firth's (2004) study suggests that major exchanges dominate the contemporaneous transfer of pricing information and also provides support for the hypothesis that markets are related through their second moments. Moreover, it indicates that information emanating from foreign markets has become more important for domestic markets and that this can be attributed to the occurrences of financial crisis.

Ferreira and Gama (2007: 3169, 3171) study the impact of sovereign debt rating news on a (noteworthy) geographically representative sample of 29 developing and developed countries. The transmission mechanism that is proposed is that ratings provide information about the future economic health of a given country and possible government policy decisions.[65] The authors report that rating changes are associated with information spillover effects; negative news (downgrades) relating to the credit outlook assigned to a specific sovereign debt country have a negative impact on the markets of the non-event countries. Information spillovers are more pronounced for emerging economies, for countries that belong to the same trade block and countries that are similar as a result of proximity. Ferreira and Gama (2007) conclude that although rating upgrades do not appear to have an impact on market returns, as they may be anticipated, rating downgrades are associated with a negative impact. This is attributed to the transmission of information to stock markets.

Singh, Kumar and Pandey (2010: 55) study information spillovers across Asian, European and North American markets.[66] The authors argue that liberalised capital markets, financial reforms and technological advances have increased the response of domestic markets to news and shocks from the rest of the world and have strengthened stock market linkages. Notably, Singh e*t al.* (2010) propose that the same information affects all markets similarly.

[65] Ratings issued by Standard and Poor (S&P). The countries are Argentina, Austria, Belgium, Brazil, Canada, Chile, China, Colombia, Denmark, Finland, Greece, Hungary, Indonesia, Ireland, Israel, Italy, Korea, Malaysia, Mexico, New Zealand, the Philippines, Poland, South Africa, Spain, Sweden, Thailand, Turkey, the UK and Venezuela. Therefore, the sample is representative of North America, South America, Europe, Asia and Africa.
[66] Singh *et al.'s* (2010: 56) study is noteworthy in its simultaneous coverage of multiple markets. The following markets are considered; China, Hong Kong, India, Indonesia, Japan, Malaysia, Pakistan, Singapore, Taiwan (Asia), Germany, France, UK (Europe), Canada and the US (North America). Together these markets span the complete time and geographical horizon.

On the basis of a VAR analysis, the authors find that Asian markets widely reflect information spillovers from multiple European (e.g. German and French) and North American markets (US and Canadian).[67] Similarly, European markets are impacted by information emanating from multiple markets; these being (other) European, North American and Asian markets. The US market responds to information originating from certain Asian (e.g. Indian, Korean) and European markets (e.g. German) and the Canadian market. Singh *et al.* (2010: 61-62) also investigate volatility spillovers, arguing that information flows across markets may have a strong impact on volatility which is regarded as a (better) proxy for information. The results of an AR(1)-GARCH(1,1) model indicate that there are significant volatility spillovers between Asian, European, US and Canadian stock markets. In summarising the findings of the study, the authors state there are information flows between markets and, similarly to Kim (2003), the US market is the most influential.

Beirne, Caporale, Schulze-Ghattas and Spagnolo (2010: 252) study information spillovers in the Middle East and Latin America and test for spillovers in returns and variances.[68] The results of a tri-variate VAR-GARCH(1.,1)-in-Mean model indicate that that regional and global stock market linkages matter for most emerging markets. This is especially applicable for Asia, Europe and Latin America. Notably, Beirne *et al.* (2010) reject the joint null hypothesis of no regional spillovers and/or no global spillovers for almost 90% of the countries in the sample. Asian markets are dominated by linkages with global markets with only China, Sri Lanka, Taiwan and Korea dominated by regional linkages. In contrast, Latin American markets are dominated by regional linkages although a number of countries exhibit both global and regional linkages. Although regional linkages dominate emerging European markets, under two thirds of the economies in the sample are also globally integrated and half are both globally and regionally integrated. Spillovers also extend to volatility. The null hypothesis of no volatility spillovers from regional and/or global markets is rejected for 85% of emerging European markets and South Africa, half of the emerging economies in Asia and Latin America and for over a quarter of emerging economies in the Middle East and North Africa. The authors conclude by stating that investors should be concerned with linkages between returns and volatility in emerging markets when

[67] For example, Singh *et al.* (2010: 60) report that the Asian markets respond to movements in the French (CAC 40), German (DAX 30), US (Nasdaq) and the Canadian (S&P/TSX 60) markets.
[68] The study encompasses 41 emerging market economies grouped according to their geographic designations, which are Asia, Latin America, Europe, the Middle East and North Africa (Beirne *et al.,* 2010: 253).

formulating investment strategies and expectations. The study shows that information spillovers are widespread, are evident across geographic locations and are inter- and intra-regional in nature. This attests to the complexity of market interdependencies and the complexity of inter-border information transmission mechanisms.

Hanousek and Kočenda (2011: 170, 176) study the impact of (foreign) news emanating from the European Union (EU) and the US on three emerging European markets which , at the time of their study, have begun integration with the European Union (EU). These are the Czech, Polish and Hungarian stock markets.[69] The potential for these markets to reflect foreign macroeconomic news announcements[70] is attributed to foreign investor presence and trade links with the EU. Using regression analysis, Hanousek and Kočenda (2011) report that all three markets respond to macroeconomic news emanating from the Eurozone and/or the US with regard to prices, the real economy and the business climate. All three markets are also integrated with their respective regional stock markets and the US and Eurozone markets.[71] Hanousek and Kočenda (2011) conclude that interactions with the developed markets are strongly determined by macroeconomic news originating from developed stock markets and this is attributed to perceptions relating to the economic climate conveyed by macroeconomic news.

The literature in this section indicates that markets reflect information that emanates from other external markets, especially those that are considered to be information leaders. Relevant information may emanate from either regional and/or international markets and is often of a macroeconomic nature. Information spillovers are reflected in both returns and volatility. That it is information that impacts markets is an important theme that runs throughout this section. While the US appears to be an information leader, other markets such as the Japanese and European markets, are also important (Kim, 2003; Wang & Firth, 2004). Therefore, it follows that information originating from information leaders, other important markets and global markets in general, should be considered in representations of the linear factor model and the APT relation.

---

[69] As measured by the respective national indices, the Prague Stock Exchange Index (PX), the Warsaw Stock Exchange Index (Wig-20) and the Budapest Stock Index (BUX).

[70] Macroeconomic announcements for the Eurozone and the US are considered in regressions whereas co-movement is accounted for by using the Dow Jones to reflect US stock market movements and the German DAX to reflect the impact of Eurozone stock markets. See Hanousek and Kočenda (2011: 172) for a full list of the macroeconomic factors considered.

[71] For example, regional integration for the Hungarian stock market is established from its relationship with the Polish and Czech stock markets.

4.3. INTERNATIONAL INFLUENCES AND THE APT

4.3.1. How The APT Approaches International Influences

The literature in Section 4.2.1. demonstrates that international markets exhibit interdependence. The literature in Section 4.2.2. shows that market movements are related to information spillovers. Therefore, markets are influenced by macroeconomic news emanating from information leaders or otherwise, which provides an explanation for market co-movement in the form of shared responses to macroeconomic news. This motivates for the consideration of global influences in the linear factor model. Fortunately, the APT already takes this aspect into account by permitting international influences to feature in the framework.

APT models and multifactor models based on the APT's linear factor model can be divided into three variants. The first variant assumes that investors are restricted to domestic assets only and that returns are solely described by domestic factors. Cauchie *et al.* (2004: 168) argue that this is a reasonable assumption if investors operate in a closed economy or if the market is completely segmented from international markets. Examples of such models are those of Chen *et al.* (1986) and Berry *et al.* (1988). The second variant relaxes the assumption of investment being purely domestic and it is assumed that markets are perfectly integrated and investors trade assets internationally. This leads to the international APT. Under this assumption, returns are assumed to be solely driven by international influences. Examples of such models are those of Ferson and Harvey (1994) and Harvey (1995). Differences between these variants are attributable to the assumptions about the levels of market integration; perfect integration or perfect segmentation represent extremes (Ferson & Harvey, 1994: 777; Bilson *et al.,* 2001: 404; Cauchie *et al.,* 2004: 168).

Finally, there is also a third variant that assumes partial integration or partial segmentation. Cauchie *et al.* (2004: 168)state that although it is reasonable that markets will be integrated in an era of increasing globalisation, there are barriers to international investments and additional costs that are associated with such investments. Therefore, an APT model that assumes partial integration may provide a better explanation of return behaviour. The third variant therefore relates returns to both international and domestic factors (Bilson *et al.,* 2001: 404). Examples of such models are those of Clare and Priestley (1998) and De Lint (2002). The flexibility of the APT permits for the combination of domestic and global factors and the incorporation of global factors in the form of a residual market factor derived from a global market index. This will fulfil the role of a second residual market factor and act as a test for omitted factors. If such a factor is significant, then this implies that all omitted factors may not be reflected by the factor set that enters the linear factor model and a conventional residual market factor. A second residual market factor will be required to reflect global influences. Section 4.3.2. focuses on the latter two variants, namely those assuming perfect integration or partial integration/segmentation, and demonstrates how the APT accommodates global influences.

4.3.2. The APT And International Influences

Ferson and Harvey (1994: 777) investigate the extent to which global factors explain movements on 18 national stock markets. Although the assumption of fully integrated markets is the basis of the modelling approach, the authors note that this is an extreme assumption. It is argued that this assumption is unlikely to provide a good approximation of the actual complexity of international investments and any results should be interpreted as a baseline case. A total of seven global factors are considered together with returns on the MSCI World Market Index, which is used to proxy for the world market portfolio. Regressions of returns indicate that the global market index is the most important factor that explains returns. The MSCI World Market Index by itself has a statistically significant impact on all markets with an average $\bar{R}^2$ of 0.337 (Ferson & Harvey, 1994: 786). The other three factors that are identified as important are the exchange rate, long-term inflation expectations and changes in the crude oil price. The combination of these factors with the global market index produces a marginal gain in explanatory power; the average $\bar{R}^2$ increases to 0.36. In cross-sectional tests, Ferson and Harvey (1994) find that although priced, the contribution of the world market index is weak. On average, the index explains 4% of the variation in expected returns and results in relatively large pricing errors. Combining the world market index with the other factors reduces pricing errors[72] and eliminates abnormal average performance for certain markets. These results indicate that global factors play a role in pricing. They also show that the world market index is important for explaining the return generating process. These results also confirm the validity of these factors as APT factors by showing that these factors play a role in explaining the cross-section of expected returns and account for risk in returns.

[72] Defined as the difference between average country returns and expected returns as predicted by cross-sectional models (Ferson & Harvey, 1994: 790).

Harvey (1995: 20-21) extends the investigation of investment risk to emerging markets arguing that the level of integration can vary across markets, depending upon development. The author argues that as the level of integration lessens (increases), the sensitivity of returns to local risk factors increases (lessens). A total of 21 developed markets and 20 emerging markets are considered and returns on national markets are related to excess returns on the MSCI World Index, a trade-weighted currency index, oil prices, OECD (Organisation for Economic Co-operation and Development) industrial production and OECD inflation. For the developed markets, 20 have a significant exposure to the international market index, eight have a significant exposure to oil, two have a significant exposure to industrial production and five have a significant exposure to inflation. For the emerging markets, six have a significant exposure to the market index, five have a significant exposure to oil, three have a significant exposure to world industrial production and four have an exposure to world inflation. The market index explains most of the variation in returns, as in Ferson and Harvey (1994). For developed markets, the average $\bar{R}^2$ increases from 0.33 (for a single-factor model) to 0.37 (when all factors are considered) and the average $\bar{R}^2$ increases from 0.03 to 0.05. As emerging economies are often characterised by a shifting industrial structure that results in changes in risk sensitivities over time, Harvey (1995: 42-45) also investigates the changing correlation between emerging market returns and returns on the MSCI World Market portfolio. Results indicate that the level of correlation has increased for a number of emerging markets, indicating increasing integration. The author concludes by stating that emerging markets are not well integrated into the global economy although there is evidence of increasing integration. Therefore, global factors are likely to be presently more important in emerging markets than in the 1990s and prior to the 1990s, given the evidence of increasing integration for emerging markets.

Kavussanos, Marcoulis and Arkoulis (2002: 924) extend the study of the role of global influences to international industry returns. It is argued that certain industries are increasingly globalised as a result of cross-border operations, mergers and alliances. The authors motivate for a multifactor specification to describe returns by referring to the APT. Returns on 38 global industries are related to returns on the MSCI World Market Index and a global factor set consisting of unexpected changes in global exchange rates, global credit risk, oil prices and industrial production. Returns on the MSCI World Market index are significant for every industry and this is also the most important factor. The average $\bar{R}^2$ increases from 0.602 to 0.614 when this factor is combined with the other global factors.

The other factors are far less important, a finding that Kavussanos *et al.* (2002) attribute to the specific characteristics of each industry.[73] The authors conclude that amongst the factors considered, the return on the world market portfolio is the most important factor, a finding similar to that of Ferson and Harvey (1994) and Harvey (1995). The relatively minor role of the other factors and the high levels of explanatory power associated with the MSCI World Market Index indicate that most of the global influences in returns can be summarised in a multifactor specification by a world market portfolio.

Choi and Rajan (1997: 31) take a somewhat different approach by proposing that markets are either partially segmented or partially integrated rather than a "polar case of complete segmentation or integration," as assumed in Ferson and Harvey (1994), Harvey (1995) and Kavussanos *et al.* (2002). The authors test a joint hypothesis of market segmentation and the presence of an exchange risk factor in the linear factor model. The IAPM, which Choi and Rajan (1997) argue is consistent with the APT (Section 2.3.2.), is used as a basis of the investigation with it being argued that such a factor structure is justified by partial segmentation arising from barriers to international capital flows. Consequently, a three-factor model is specified to test the joint hypothesis set out above, incorporating the exchange rate and two factors, the respective domestic and international indices. These indices are the respective national indices for the markets in the sample and the MSCI World Market Index. The Canadian, French, Japanese and the UK stock markets are found to be partially segmented; both the domestic and world market indices are priced showing that expected returns are explained by domestic and global influences as opposed to purely domestic or global influences. Germany is found to be fully segmented, expected returns are explained only by the domestic index; Italy is found to be fully integrated, as evident from the significance of the risk premium on the global market index. The risk premium on the exchange rate is significant for France, Germany, Italy, Switzerland and the UK although the respective risk premia have different signs. The results of the linear factor model indicate that domestic market returns are important for stocks in all markets and international risk is important for France, Germany and Japan suggesting that these three markets are partially integrated. Choi and Rajan (1997) conclude that national markets can be described as partially segmented and partially integrated rather than polar opposites of full integration or

[73] For example, oil is significant for just over a quarter of the industries in the sample and global inflation, the next most important factor, is significant for just over a fifth of the sample (Kavussanos *et al.,* 2002: 929).

full segmentation. This suggests that there is a role for both domestic and global factors in the linear factor model and the broader APT context.

Bilson *et al.* (2001: 403-404) acknowledge that perfect integration nor perfect segmentation are realistic assumptions and therefore both domestic factors and global factors may be important in determining return behaviour. The authors investigate the relationship between macroeconomic factors and emerging markets, arguing that emerging markets go through stages of emergence. This suggests that phases of higher segmentation will be associated with the increased importance of domestic factors. To investigate the ability of local and global factors to explain return variation, Bilson *et al* (2001: 405) refer to the APT to motivate for a multifactor specification that incorporates both domestic factors and a global market index, namely the MSCI World Market Index. Regression results for 20 emerging markets indicate that half of the markets are influenced by movements in the MSCI World Market Index. This suggests that emerging markets generally show low levels of capital market integration. Other factors that are influential are the exchange rate and the money supply whereas prices and real activity are significant for a single market (Mexico and Portugal). On the basis of poor support for the relevance of the factors considered and generally low $\bar{R}^2$ values (some as low as 0), the specification is extended by including a number of other domestic and microeconomic factors and notably, regional indices. Explanatory power increases substantially; the average $\bar{R}^2$ is now equal to 0.60 and the regional market indices are significant for over half of the markets in the sample. Although not fully conclusive, these results indicate that global and regional market indices contribute to explaining returns. The findings of Bilson *et al.* (2001) provide support for partial integration/segmentation that is also of a regional character.

De Lint (2002: 60-62) compares the importance of domestic and global factors in the Mexican stock market and six Asian markets. A critique of the assumption of full market integration is offered with it being argued that under the assumption of full integration only global factors are considered and these are unlikely to account for local crises. Moreover, the assumption of perfect integration for emerging markets is credited with observed rejections of the asset pricing models in these markets. A proposed solution is the inclusion of both global and local factors in specifications that permit global factors to show their effect on these markets and allow for a comparison of the respective importance of these factors. To investigate the importance of global and domestic factors, De Lint (2002) refers to the APT and specifies a multifactor model incorporating five global factors (G7 inflation, industrial production, long-term interest rates, short-term interest rates and the MSCI World Market Index) and three domestic factors (inflation, industrial production and the local currency to the Dollar exchange rate). Global factors are found to be important but are not the only factors that feature in the return generating process. Domestic factors are still found to be important sources of systematic risk. Significant exposures to these factors are contrary to full integration. As in the other studies, the world market index is the most important factor, followed by G7 short-term interest rates and G7 industrial production. Only Malaysia is not influenced by the world market index. The most important local factor is the exchange rate followed by inflation and industrial production. Next, De Lint (2002) goes onto investigate the changing importance of local and global factors over time for countries that went through a crisis. For the Philippines, Singapore, Mexico, Korea and Thailand, local factors become important prior to a given crisis and remain important during and after the crisis. Therefore during times of stability, investors are more concerned about global factors but during and around the time of a crisis, local factors become more important. The author concludes that the markets in the sample are mostly impacted by global factors, especially by returns on the world market index. Local factors are nevertheless also important, especially around times of crises and the results are indicative of partial integration. Importantly, this study, as the studies of Choi and Rajan (1997) and Bilson *et al.* (2001), provides support for linear factor model specifications that combine both domestic and global factors.

This section focuses on literature that assumes either full integration or partial integration or partial segmentation. All studies discussed are underpinned by the APT and specifications are motivated by the linear factor model that is the basis of the APT. These studies also demonstrate how the APT incorporates global influences. Full integration appears to be an extreme. As suggested by Choi and Rajan (1997), Bilson *et al.* (2001) and de Lint (2002), returns can be described by a mixture of domestic and global factors. Moreover, in studies that assume full integration, the MSCI World Market Index is the most important factor. Although the findings at times are somewhat mixed, especially when emerging markets are considered, it is almost certain that integration has increased for these markets. The influence of global market factors will therefore be more relevant than ever. Cauchie *et al.* (2004: 168) state that typically, a world market portfolio is considered as a source of global risk. The literature in this section shows that a world market portfolio, in the form of the MSCI World Market Index, is generally important for returns regardless of whether full or partial integration is assumed. Moreover, the MSCI World Market Index features in all of these studies suggesting that this is a viable candidate factor for a second residual market factor under the seemingly more realistic assumption of partial integration/segmentation.

4.4. INFORMATIONAL CONTENT

Much like the conventional residual market factor is proxy for omitted factors, a world market index and the derived residual market factor may be considered as a proxy for global influences (Abugri, 2008: 397, 400). Van Rensburg (1995) and Szczygielski and Chipeta (2015) propose that the South African stock market is integrated with global markets. Given the prominent role of global factors in the APT and stock markets in general, it makes sense to incorporate a world market index into the return generating process and to use this factor to determine the adequacy of a conventional residual market factor derived from the domestic market index. As in Section 3.5., the informational content of a world market index is briefly explored with the assumption being that correlation is indicative of information reflected in an international market index.

Mateus (2004: 242) investigates the risk and predictability of returns for 13 EU accession countries using global factors. The global factors selected are those widely used in the literature, namely changes in 90-day EU-US treasury yield spreads, fluctuations in the Dollar to global major currency exchange rate, unexpected G-7 monthly inflation, changes in expected inflation, world oil prices and changes in G7 industrial production. In preliminary analysis, Mateus (2004) reports that a general European market aggregate, the MSCI Europe Index, is significantly correlated with returns on the MSCI World Market Index. Some of the market indices in the sample are also correlated with the MSCI World Market Index. This suggests, depending upon the direction of information flows, that a global market index may reflect information originating from regional markets as well as external markets. Notably, the reported high level of correlation (0.91) between the MSCI World Market Index and the European market aggregate implies that the MSCI World Market Index can be a proxy for regional influences. Furthermore, returns on the MSCI World Market Index are correlated with changes in the exchange rate, changes in expected inflation, changes in G7 industrial production and the EU-US yield spread suggesting that the index is a proxy for these global factors. Brown *et al.* (2009: 302), in their study of risk premia in international markets (Section 3.4.), derive an international residual market factor by regressing returns on the MSCI World Market Index onto a set of global macroeconomic factors. The representation of this global return generating process shows that returns on this index are significantly related to US, Japanese and Euro bond yield spreads, changes in the Yen-Dollar exchange rates and a US Fama–French HML factor, which in itself reflects macroeconomic information (also see Aretz, Bartram & Pope, 2010). Together, these factors (and other factors in the specification) explain 23.76% of variation in global returns. This suggests that the MSCI World Market Index reflects economic conditions in three large and important economic regions, namely Europe, Japan and the US, which would otherwise be measured by individual factors in the form of yield spreads, the Yen-Dollar exchange rate and the HML factor.

Moerman and Van Dijk (2010: 844) investigate the pricing of inflation risk in international asset returns. Preliminary analysis indicates that the MSCI World Index is strongly correlated with the real and nominal exchange rates for France, Japan, the UK and US and the inflation differential for the US These factors are impacted by macroeconomic announcements themselves, suggesting that the MSCI World Market Index summarises information reflected by these factors and secondary information that is reflected in these factors. It follows that there is a hierarchical flow of information and this is aptly demonstrated by Andersen, Bollerslev, Diebold and Vega (2003: 38; 51), who undertake a study of the impact of macroeconomic news on exchange rates. The impact of US and German macroeconomic news releases is considered for the German Mark, British Pound, Japanese Yen, Swiss Franc and the Euro/Dollar exchange rates. The authors find that exchange rates respond to US announcements relating to real activity, consumption, investment and government expenditures, cyclical, inflation and monetary policy. A similar argument could be made for inflation (and inflation differentials) which will partially reflect the underlying macroeconomic fundamentals. As Moerman and Van Dijk (2010) consider exchange rates and inflation differentials for four major global economies, it follows that changes in macroeconomic conditions in these economies will impact other integrated markets, as suggested by Andersen *et al.* (2003). The high correlation of these factors with returns on the MSCI World Market Index indicates that the inclusion of this index will transmit information related to changes in macroeconomic fundamentals in these important economies that are associated with factors that are not directly considered. Furthermore, Moerman and Van Dijk (2010: 844) report that the MSCI World Market Index is highly correlated with returns on the MSCI indices for Germany, France, Japan, the UK and the US and is most highly correlated with the US, followed by the UK and Japanese markets. This suggests, as shown in the literature in Section 4.2.2., that a world market index reflects information spillovers from these markets, regional markets and information emanating from information leaders.

The literature indicates that much like the conventional residual market factor, a residual market factor derived from a world market index will proxy for global macroeconomic information. Such a proxy will reflect information emanating from information leaders, notably the US, regional influences and information relating to economic conditions prevailing in major markets and global economic conditions in general. Therefore, a world market index, in the form of the widely used MSCI World Market Index, is an apt candidate for a second residual market factor. The inclusion of the international residual market factor as a second proxy for omitted influences will constitutes a test of the adequacy of the residual market factor. If the conventional residual market factor is an adequate proxy for all remaining factors, then a second residual market factor will be irrelevant.

4.5. MOTIVATION FOR A SPECIFIC MARKET INDEX

Before concluding this chapter, the reason for considering the MSCI World Market Index for the derivation of a second residual market factor, aside from widespread application, deserve further mention. Born and Moser (1988: 289) state that the formation of the *true* market portfolio is an aggregation process that incorporates all influences underlying the factors in the return generating process. Brown and Brown (1987: 31) state that a market index should reflect the relevant universe of assets but argues that “the quest for an all-inclusive ‘market portfolio’ is, from a practical standpoint, a futile one.” It follows that the MSCI World Market Index is unlikely to be the true global market portfolio (much like the JSE All Share Index in the South African context). Nevertheless, Harvey (1995: 28) recognises that although the MSCI World Market Index does not reflect investment in emerging markets, it is the most widely used world benchmark.

It can be argued that this specific index, the MSCI World Market Index, should not be used to derive the second residual market factor as it is not sufficiently inclusive. Brown and Brown (1987:31) argue that the solution is to define the market in terms of its relevant components (relevant to the sample). Later on, this study nevertheless proceeds to use the domestic market aggregate and the MSCI World Market Index to derive the residual market factors. This is motivated, partially, by the widespread use of a domestic market aggregate and the MSCI World Market Index to account for omitted factors. Custom indices could be constructed to capture omitted influences and to determine whether any influences have not been captured by the domestic market index. Nevertheless, a solution to the omitted factor problem must be practicable and easily implementable. As evident from the literature, the use of a residual market factor derived from a national market aggregate and the use of the MSCI World Market Index to proxy for international influences is widespread.

4.6. CHAPTER SUMMARY AND CONCLUSION

The aim of this chapter is to investigate the role of global influences and factors in returns. The literature in Section 4.2.1. shows that markets exhibit interdependence and that interdependence has grown over time and following financial crisis. Intra-regional correlations also indicate that regional and not only global developments have an important influence on markets. The complexity of these interdependencies is compounded by multidirectional information flows. Interdependence is driven by greater liberalisation, improvements in technology, lower financial transaction costs and cross-border operations. Section 4.2.2. suggests that these interdependencies and co-movements are driven by simultaneous responses to macroeconomic information originating from major developed markets, regional markets and information leaders. Linear factors models should therefore incorporate global influences.

Literature on the APT acknowledges that global influences can enter the linear factor model and the pricing relation. Extensions of the APT assume either partial market segmentation/integration or full integration (Section 4.3.1.). What emerges from the literature is that a world market index is an important factor in returns. A world market index can be combined with domestic factors if partial market segmentation/integration is assumed to capture the influence of global factors (Section 4.3.2.). Such an index will reflect the direct and indirect influence of other global factors and global and regional market movements that are driven by macroeconomic information and information spillovers (Section 4.4.). This provides support for a second residual market factor within the linear factor model that is derived from the widely used MSCI World Market Index. The widespread use of this index offers a practicable and readily implementable solution for incorporating global influences into the linear factor model (Section 4.5.).

The inclusion of a second residual market factor will constitute a test of the adequacy of the conventional residual market factor. It follows that if combined with the conventional residual market factor and other factors in the linear factor model, all systematic variation in returns should be exhausted in the return generating process if underspecification is not eliminated in the first instance by the conventional residual market factor. If the residual market factor is not an adequate proxy, then the value of incorporating a second residual market factor may be assessed and a pronouncement on the efficacy of a two residual market factor approach can be made.

CHAPTER 5

UNDERSPECIFICATION AND THE ARBITRAGE PRICING THEORY

5.1. INTRODUCTION

The aim of this chapter is to set out the consequences of factor omission within the context of the APT. Literature motivated by Chen *et al.'s* (1986) macroeconomic APT relies upon macroeconomic factors to describe returns and resolves potential underspecification by incorporating a residual market factor. This residual market factor, derived from a broad domestic market aggregate, is hypothesised to be a catch-all proxy for omitted and unobservable factors (Section 3.2.). The literature also indicates that more than one residual market factor may be necessary to account for omitted factors (Section 3.4.). In order to assess the efficacy of the residual market in resolving underspecification and the adequacy of the macroeconomic linear factor model, it is necessary to set out and outline the consequences of underspecification. This chapter therefore defines underspecification and discusses its impact on model estimation and interpretation. As this study is concerned with the linear factor model motivated by the APT, the discussion is framed within the context of the APT.

Section 5.2. proceeds by defining underspecification and providing reasons as to why underspecification can plausibly and readily arise when specifying a macroeconomic linear factor model. Section 5.3. sets out the econometric consequences of underspecification, which are illustrated with reference to a simulation exercise. This demonstrates the immediate and unsubtle impact of factor omission on model estimation and interpretation. Section 5.4. discusses the consequences of underspecification. The discussion is framed within the context of the immediate assumptions that underlie the linear factor model and extends to the broader application of the APT. This section aims to outline the more subtle consequences of underspecification and emphasises the importance of underspecification in the APT. Section 5.5. summarises and concludes the chapter.

5.2. UNDERSPECIFICATION

Model underspecification can be viewed as a form of specification error or misspecification. Gujarati and Porter (2009: 219, 221) state that specification errors occur when a model other than the correct model is estimated. The authors list the following as types of misspecification:[74]

1) The inclusion of an unnecessary factor(s).
2) The adoption of the wrong functional form.
3) Errors in measurement in the dependant factor.
4) The omission of relevant factor(s).

This study is concerned with the last type, the omission of relevant factor(s), namely underspecification, and the ability of the residual market factor to resolve underspecification. Reasons for the purposeful or accidental omission of relevant factors from the linear factor model and therefore the APT model are numerous and follow intuitively (Gujarati, 2004: 45-46):

1) The vagueness of theory
2) Unavailability of data
3) Core factors versus peripheral factors
4) Intrinsic randomness of human behaviour
5) Poor proxy factors
6) Principle of parsimony
7) Wrong functional form

Although this list is extensive, certain reasons are more relevant to the APT than others. For example, Panetta (2002: 421) states that the APT does not indicate which specific factors should enter the linear factor model, systematic factors in the APT model are not explicitly identified and the linear relationship between returns and factors is an assumption of the framework. However, the identification of factors can be guided by reference to the (widely used) dividend discount model or by using factor analysis (Azeez & Yonezawa, 2006: 577; Yao *et al.,* 2014: 945). Similarly Bilson *et al.* (2001: 405) note that the selection of factors is subject to criticism as it is subjective and arbitrary in nature. The authors argue that the selection of factors can be accomplished through reference to prior research and judgement. Panetta (2002) and Bilson *et al.* (2001) highlight just some of the challenges of

[74] Gujarati (2004: 509) makes a distinction between specification errors and model misspecification errors. Underspecification is considered to be a form of misspecification. A model specification error occurs if the true model specification is known but is not estimated. Model misspecification occurs when the true structure of the specification is not known. This is particularly relevant to the linear factor model as the true structure of the linear factor model is unknown. This study avoids making an explicit distinction and refers to both types of errors as misspecification.

selecting macroeconomic factors that will explain return behaviour. There may be disagreement on the appropriate approach to selecting factors. This suggests that identifying a relevant and sufficiently exhaustive macroeconomic factor set that adequately describes the true return generating process is not straightforward.

Van Rensburg (1995: 48; 1996: 106) states that while factors for which timely and accurate data is readily available should be selected, certain types of data, such as (South African) corporate bond data (at the time of the author's study), is considered unreliable. Other data, such as the personal savings rate, is subject to constant revisions and inaccuracies. Clements and Galvão (2008: 546) state that the unavailability of some types of data at higher frequencies, such as data for GNP and GDP, which is usually available at quarterly frequencies, forces specifications to be estimated at lower frequencies. This suggests that at times, for example, when monthly data is required, data for important and relevant factors may not be available at the desired frequency. Van Rensburg (1996: 106), in the study of priced APT factors on the JSE, uses monthly data and notes that GDP data is only available quarterly. Therefore, although this factor may be relevant and important, it is omitted from the analysis as a result of the unavailability of data at the desired frequency. This is further compounded by the unobservability of some factors and therefore, the unavailability of data for such factors in the first instance (Sykes, 1993: 27).

APT studies often include the most important factors in the linear factor model, but omit factors that are perceived to be less important. Jointly, the included factors should capture as much of the systematic variation attributable to changing economic conditions as possible (McElroy & Burmeister, 1988: 41). However, it is plausible that some peripheral yet important factors will be omitted. Hughes (1984: 207) reports that although 12 statistically derived factors are sufficient to explain half of the variation in returns on stocks on the Toronto Stock Exchange, the first factor accounts for almost a third of the variation in returns and the remaining 11 factors account for approximately a fifth of the variation. The first factor appears to be a core factor. Although the remaining factors are not important individually, jointly, these factors make a significant contribution to the explanation of returns yet may be excluded in analysis. Therefore, an omission of factors that appear peripheral and trivial individually, but together contribute substantially to explaining returns will result in underspecification. Moreover, there is nothing that precludes seemingly trivial factors from being systematic in nature.

The APT framework relies on pervasive factors, as proxied by macroeconomic factors, to explain returns. It is nevertheless possible that sentiment and behavioural factors and not only macroeconomic factors impact stock price movements. Also, investors may not be rational and will therefore respond irrationally (or will not respond at all) to macroeconomic news (Malkiel, 2003). In a study of the determinants of REIT returns, Lin, Rahman and Yung (2009: 460) find that investor sentiment[75] is more important relative to macroeconomic factors. In regressions of REIT returns onto a measure of investor sentiment and a set of macroeconomic factors, the authors find that investor sentiment subsumes the explanatory power of the default and term structure factors. In conclusion, Lin *et al*. (2009) state that behavioural finance proposes the use of behavioural factors in asset pricing and that the importance of macroeconomic factors has waned over time implying that the role of behavioural factors may have increased over time. This suggests that the intrinsic randomness in human behaviour may also be relevant for asset pricing and explaining return behaviour, although it may not be reflected in macroeconomic factors that enter the linear factor model.

The principle of parsimony should be applied within the APT framework. If a model requires a large number of explanatory factors, then it fails to simplify the return generating process and is therefore of questionable value. The application of the principle of parsimony suggests that the number of factors in the linear factor model should be restricted and a simpler explanation is preferable to a more complex one (Driffill, 2011: 28). Additionally, the APT framework does not specify an upper bound of factors and the number of factors may increase as the number of stocks in a group (portfolio) increases (Dhrymes *et al.,* 1984: 339). Middleton and Satchell (2001: 506) argue that if there is uncertainty relating to the number of factors in a model and there is a desire for the APT model to hold, the principle of parsimony is inappropriate. Therefore, while the principle of parsimony represents a perhaps useful and convenient abstraction of the linear factor model, it does not preclude the omission of relevant factors and resultant underspecification.

Finally, it is possible that the process describing returns is non-linear while the APT framework assumes that it is linear. Funke and Matsuda (2006: 193) study the impact of macroeconomic news on the US and German stock markets and postulate that the impact of innovations differs according to the state of the economy. Although the German stock

[75] As measured by fluctuations in closed-end fund discounts (Lin *et al.,* 2009: 454).

market does not show evidence of an asymmetric response to innovations, the same cannot be said about the US stock market during booms and recessions. The impact of news relating to GDP growth, unemployment and Federal Reserve target rates differs in magnitude and significance across economic states. For example, Funke and Matsuda (2006) report that positive changes in the target rate in a recessionary environment have a positive and significant impact on S&P 500 Index returns. In contrast, the impact is negative and inconsistent in significance during high and medium growth environments. Reinganum (1981: 320), in an investigation into the validity of the APT, recognises that the return generating process may not be linear and therefore is not described by the linear factor model. If this is the case, as would be under the purest interpretation of the linear factor model implied by APT theory, the linear factor model is underspecified as it fails to explicitly account for asymmetry and non-linearity in returns. It is of the incorrect functional form.

5.3. THE ECONOMETRICS OF UNDERSPECIFICATION

5.3.1. Econometric Consequences

This section provides an overview of the econometric consequences of model underspecification. For the purposes of this summary and simplifying abstraction, the consequences of underspecification are discussed in relation to the least squares framework.[76]

A hypothetical linear factor model specification that is assumed to represent the true return generating process can be represented as follows (Clarke, 2005: 342):

$$R_{it} = \alpha + b_1 f_{1t} + b_2 f_{2t} + b_3 f_{3t} + \varepsilon_{it} \tag{5.1}$$

where R_{it} is the return on stock *i* at time *t* and f_{1t}, f_{2t} and f_{3t} are the factors that adequately explain the return generating process. However, owing to the hypothetical unavailability of data, the exclusion of peripheral but relevant factors, the intrinsic randomness of human behaviour not captured in observable or quantifiable factors, the use of poor proxy factors

[76] The least squares methodology is widely applied, easily understood and least squares estimators have the theoretically desirable properties of being best linear unbiased estimators (BLUE) of model parameters. The overview provided here is based upon the least squares methodology owing to widespread application, popularity and simplicity. However, depending upon the characteristics of the data, it may be desirable to use alternative estimation methodologies, such as maximum likelihood (ML) estimation. ML estimators will have BLUE properties if the residuals are normally distributed (Smith & Hall, 1972: 186). Therefore, it is postulated that this overview is instructive and sufficiently general and can be extended to other regression methodologies.

and the (potentially misguided) principle of parsimony (Section 5.2.), the following specification is estimated, omitting f_{3t}:

$$R_{it} = \gamma + \alpha_1 f_{1t} + \alpha_2 f_{2t} + \upsilon_{it} \tag{5.2}$$

The consequences of omitting f_{3t}, under the ordinary least squares (LS) methodology can be summarised as follows (Lehmann, 1990: 72; Dominguez, 1992: 96; Sykes, 1993: 25-26; Van Rensburg, 2002: 91; Sadorsky & Henriques, 2001; 204; Brauer & Gómez-Sorzano, 2004: 38; Gujarati & Porter, 2009: 222; Bucevska, 2011: 631; Studenmund, 2014: 179-180):

1) If f_{3t} is correlated with the remaining factors, f_{1t} and f_{2t}, and the correlation between these factors is non-zero, then the intercept in equation (5.2), γ, and corresponding coefficient estimates, α_1 and α_2 will be biased and inconsistent. The estimated intercept and coefficients will not be equal to their true values and they will remain biased even if the sample size increases infinitely. In other words, γ, α_1 and α_2 are under- or overestimated and the extent of the bias is dependent upon the correlation between the omitted factor and included factors. If f_{3t}, the omitted factor, has a positive (negative) impact on R_{it} (the dependent factor) and is positively (negatively) correlated with f_{1t} and f_{2t}, then α_1 and α_2 will overestimate (underestimate) the true b_1 and/or b_2. In such a situation, f_{1t} and/or f_{2t} are credited for the indirect influence of f_{3t} as the influence of this factor is reflected in the respective coefficients of f_{1t} and/or f_{2t}.
2) If f_{1t}, f_{2t} and/or f_{3t} are uncorrelated, the intercept in equation (5.2), γ, will be biased but α_1 and α_2 will now be unbiased and consistent, unless the impact of the omitted factor on returns, f_{3t}, is zero. The bias in γ will arise as the intercept will reflect part of the impact of the omitted factor.
3) The estimated variance of the residual terms is biased. The estimated variance of the residuals in the correctly specified model in equation (5.1), ε_{it}, differs from the estimated variance of the residuals of the underspecified model in equation (5.2), υ_{it}

. The impact of the omitted factor will be relegated to the residuals and the residual variance will now also reflect dispersion that is associated with the omitted factor.

4) Conventional estimates of the variance of γ, α_1 and α_2 will not be efficient and will be biased estimates of the true variance. This will hold true even if f_{1t}, f_{2t} and/or f_{3t}, are uncorrelated. The estimated variance of the coefficients will be over-estimated and consequently, the standard errors (which measure the precision/accuracy of estimated coefficients) will be overstated. This will result in unnecessarily large confidence intervals and hypothesis tests will yield misleading inferences about statistical significance. Notably, the hypothesis that the true value of an estimated coefficient is zero (or any other null hypothesis) will be accepted more often.
5) Predictions based upon the underspecified model will be unreliable and inaccurate.
6) The omission of relevant factors may result in the residuals appearing to have non-constant variance attributable to heteroscedasticity and/or serial correlation that are impure in nature.

The summary above indicates that underspecification can have a potentially adverse impact on the estimation of the linear factor model, the APT model and inference making in general. An example that considers the less subtle consequences of underspecification is outlined next.

5.3.2. An Illustrative Example

The example that follows is based upon a simulation from Sykes (1993: 13). Although this is not an example derived from the field of finance and is not time series orientated, it is sufficiently demonstrative and provides an overview of the immediate and unsubtle effects of underspecification. Therefore, parts of it are restated below for demonstrative purposes. This hypothetical example centres on gender discrimination in the workplace. The question posed is whether women earn less after controlling for factors that can permissibly justify a lower income, namely schooling, aptitude and experience, have been taken into account. A hypothetical true model is proposed:[77]

$$\begin{aligned} & \mathit{Earnings} = 5000 + 1000\mathit{School} + 50\mathit{Aptitude} + 300\mathit{Experience} \\ & -2000\mathit{Gendum} + \mathit{Noise} \end{aligned} \tag{5.3}$$

[77] The specification is restated identically as in Sykes (1993: 13). Data is simulated by the author for the purposes of this example.

where Gendum is a gender dummy factor that equals one (1) for women and zero (0) for men, and quantifies the impact of gender on earnings. School is the number of years of schooling, Aptitude is the score from an aptitude test and Experience is the number of years of experience in the work force. Generated noise terms (residuals) are independent and their expected value is zero and coefficient estimates are unbiased, consistent and efficient. The model is then formally estimated using simulated data, and the following results are (faithfully) reproduced from Sykes (1993: 16):

Table 5.1: Noise Term With Standard Deviation Of 1000

Variable	"True value"	Estimated value	Standard error	*t*-statistic	Prob (2-tail)
Constant	5000.0	4784.2	945.4	5.060	.000
School	1000.0	1146.2	72.0	15.913	.000
Aptitude	50.0	39.1	6.8	5.741	.000
Experience	300.0	285.4	20.2	14.131	.000
Gendum	-2000.0	-1867.6	350.5	-5.328	.000
$R^2 = 0.964$					

Source: Sykes (1993)

The results in Table 5.1. indicate that gender impacts earnings, as evident from a significant *t*-statistic, after accounting for schooling, aptitude and experience, which are all statistically significant.[78] According to the model, women earn an estimated $1867.6 less relative to men. Sykes (1993: 26) then omits the School factor from the model. The results of the purposefully underspecified model are now as follows:

Table 5.2: Omitted Variable Illustration

Variable	"True value"	Estimated value	Standard error	*t*-statistic	Prob (2-tail)
Constant	5000.0	9806.5	4654.8	2.107	.041
School	1000.0	Omitted	-	-	-
Aptitude	50.0	107.5	25.6	4.173	.000
Experience	300.0	256.9	103.3	2.487	.017
Gendum	-2000.0	-2445.5	1779.0	-1.375	.176
$R^2 = 0.408$					

Source: Sykes (1993)

[78] Sykes (1993: 15) also simulates a model for which the residuals have a standard deviation of 3000 as opposed to having a simulated standard deviation of 1000 as in Table 5.1. The results of this model show the direct impact of an inflated variance; only school and experience have a statistically significant impact on earnings, which is in contrast to the results in Table 5.1. If the residuals were zero for every observation, the variance would also be zero, and the estimated model parameters would be the true model parameters (equation (5.3.)).

The first noticeable impact of underspecification is that the R^2 decreases from 0.964 to 0.408. This is expected; an important factor is omitted from the model. Consequently, the explanatory power of the model decreases. The second readily noticeable impact is on the estimated intercept, which more than doubles in size from 4784.2 in Table 5.1. to 9806.5 in Table 5.2. This is expected; the intercept is now biased upwards as the impact of the omitted factor is now reflected in the intercept (consequence 2) in Section 5.3.1.). The third readily noticeable impact is that the estimated coefficients are now further away from the true hypothesised model parameters; they are biased (consequence 1) in Section 5.3.1.). For example, whereas the estimated coefficient on the aptitude factor in Table 5.1. is 39.1, it is now 107.5 in the underspecified model in Table 5.2. whereas the true coefficient is 50.00. Sykes (1993) attributes this to the high (deliberately simulated) correlation of 0.69 between schooling (the omitted factor) and aptitude. Aptitude now incorrectly reflects some of the (positive) impact of schooling. The result is that the estimated coefficient for aptitude is far above its true value (Sykes, 1993: 25).

Next, the *t*-statistics have all decreased in the underspecified model. This is attributable to an increase in the standard errors (consequence 3) in Section 5.3.1.) arising from an upward bias in the residual variance associated with factor omission (Lehmann, 1990: 72). For example, the standard error for aptitude in the correctly specified model in Table 5.1. is 6.8 whereas it is 25.6 in Table 5.2. As the *t*-statistic is calculated by dividing the estimated coefficient by the standard error, a non-proportional increase in the standard error relative to the coefficient may result in a perceivable change in statistical significance. For example, whereas the coefficient on Gendum increases by a factor of 1.31 (-2445.5/-1867.6), the associated standard error increases by a factor of 5.08 (1779.0/350.5). This large increase renders this factor statistically insignificant. This erroneously indicates that gender does not have an impact on earnings and is in contrast to the correct model in Table 5.1., where gender does have an impact on earnings. From this, it is evident that underspecification impacts the magnitude of estimated coefficients and their standard errors. This results in misleading inferences. Most importantly, in this specific example, underspecification negates the main aim of the investigation of whether gender has an impact on earnings. Similarly, underspecification of the linear factor model can negate the main aim of estimating the linear factor model; that of identifying significant factors that drive returns, deriving the betas and then using these in tests of the APT relation to determine pricing.

The next section considers the impact of underspecification in the context of asset pricing and also highlights some of the more subtle effects of underspecification not demonstrated in this illustrative example.

5.4. UNDERSPECIFICATION AND ASSET PRICING

5.4.1. Impact On Assumptions

The immediate impact of underspecification, within the context of the APT, is on the underlying assumptions of the linear factor model. These are that the covariance (alternatively correlation) between the respective residuals of the linear factor model is zero (equation (2.2.)) and that the covariance (alternatively correlation) between the residuals and factors included in the specification is zero (equation ((2.3)).

The first assumption that is impacted is that of uncorrelated residuals, the diagonality assumption (as represented by equation (2.2)). Van Rensburg (2000: 36) states that this assumption is likely to be violated in specifications of the linear factor model that employ pre-specified macroeconomic factors to explain returns, as in Chen *et al.* (1986). Elton *et al.* (2014: 157) state that the validity of this assumption is determined by the appropriateness of factors in a specification of the linear factor model. Van Rensburg (1997: 63), in a study of APT factors in the (segmented) JSE, shows that the assumption of uncorrelated residuals is violated if pre-specified macroeconomic factors are used. Returns on individual South African stocks[79] are regressed onto three factors, namely unanticipated returns on the DJIA, unexpected changes in inflation expectations and unanticipated changes in the term structure of interest rates. Two factors that explain over 40% of the variation in the residuals are extracted from the resultant residual correlation matrix. This indicates that important common (systematic) factors remain in the residuals and that the model is underspecified. The extraction of these two factors and the scree plot of eigenvalues reported in Van Rensburg (1997) confirm the presence of significant pairwise residual correlation and the associated violation of the diagonality assumption. Van Rensburg (2002: 97) emphasises the importance of this assumption by stating that "while contemporary econometricians take considerable care to adjust for time series correlation in their regression residuals, the possibility of the presence of cross-sectional correlations being present are all but ignored.

[79] Van Rensburg's (1997) sample comprises 72 "well-traded" JSE stocks over the January 1980 to December 1989 period.

The presence of such correlations results in an omitted variable bias and with it a violation of one of the assumptions of the Gauss-Markov theorem."

Studenmund (2014: 101, 179), in a general discussion of model underspecification, states that if relevant factors are omitted from a given specification, the residual terms will be correlated with these factors. Moreover, if the included factors are correlated with the omitted factors, then the included factors and residuals will not be independent as omitted factors are a major component of the residuals. This is a direct violation of the second assumption, that of uncorrelated residuals and factors, underlying the linear factor model (equation (2.3)). The violation of this assumption results in endogeneity, which according to Roberts and Whited (2013: 494), leads to biased and inconsistent parameter estimates that make reliable inference making almost impossible In fact, the authors argue, that endogeneity may be severe enough to reverse qualitative inferences drawn upon the basis of a specification. Three potential causes of endogeneity are listed, namely simultaneity, measurement error and omitted factors. The last cause is of relevance to this study. Chenhall and Moers (2007: 177, 179-180) argue that the potential for endogeneity exists in virtually all studies in accounting and finance and studies that involve macroeconomic factors. The authors state that in the presence of endogeneity, one can no longer be confident that the results of a regression support the causality implied by an equation. This is because the sensitivity (coefficient) associated with an endogenous factor will be a biased estimate that reflects factors that are relegated to the residual term. The difficulty in interpreting this relationship lies in that the endogenous factor now reflects the influence of the omitted factor or factors. This has the potential to result in an erroneous rejection of the null hypothesis of no impact, as the estimated significant impact is the result of an omitted factor and not the endogenous factor.

This section sets out the immediate consequences of underspecification in relation to the two assumptions that underlie the linear factor model. The violation of these assumptions translates into an impact on inference making and coefficient estimates. The discussion that follows delves into the (at times more subtle) effects of underspecification that follow the violation of these two assumptions.

5.4.2. Impact On Application

Elton and Gruber (1988: 28, 31), in their investigation of the return generating process of stocks comprising the Nomura World Country Index (NRI 400), extract factors from four subsamples of the constituents of this index. It argued that if the extracted factors are common, then they should be interchangeable. This is supported by an examination of the correlations between the extracted factors which indicates that corresponding factors from each subsample are highly correlated across groups. As four sets of factors are extracted, it still remains necessary to determine which four factor solution is optimal. The authors examine residual correlation to determine the best factor solution from the four sets of factors. Returns on 20 portfolios that are formed from NRI 400 stocks are regressed onto each four factor solution and the resultant pairwise residual correlation matrices are examined to determine which factor solution produces residual correlations closest to zero. On the basis of this, an optimal factor solution is chosen. In taking this approach, Elton and Gruber (1988) acknowledge the importance of the assumption of uncorrelated linear factor model residuals and postulate that the most optimal factor solution should minimise pairwise residual correlations. This is expected if a given factor solution accounts for most co-movement in returns (Elton *et al.*, 2014: 157).

Lehmann and Modest (1987: 244; 259) consider whether a differing number of factors impacts alphas. Using monthly returns for mutual funds over a 15 year period,[80] mean intercepts (the alphas), indicative of superior (or inferior performance) for five, 10 and 15 factor structures are considered. Differences in the mean intercepts between the five and 10 factor structures are -2 and -87 basis points for the first two subperiods and +233 basis points for the last subperiod. Differences are almost indiscernible between mean intercepts for the 10- and 15-factor specifications. The respective differences are +11, -32 and -8 basis points. Lehmann and Modest (1987) state that with the exception of the differences in the mean intercepts estimated from five and 10 factors in the second and third subsamples, differences are small. Although the results are not conclusive, it appears that factor omission may have a significant impact on performance measures. Specifically, it may impact the estimated alphas. It is interesting to note that the differences in alphas between the 10 and 15 factor structures are lower than those of the five and 10 factor structures. This implies that the 10 and 15 factor structures are more appropriate and that a five factor structure is underspecified with omitted factors reflected in the intercepts. Importantly, these

[80] The sample period spans the period between January 1968 and December 1982 and is subdivided into three subperiods, January 1968 to December 1972, January 1973 to December 1977 and January 1978 to December 1982.

results indicate that factor omission may impact inferences relating to performance (Ferson & Harvey, 1994: 792).

Chang (1991: 387) compares the intertemporal and cross-sectional predictive ability of a linear factor model that incorporates only macroeconomic factors and a specification that includes these macroeconomic factors and a residual market factor. A comparison of the mean errors, namely the residuals of the respective linear factor models, indicates that the (purely) macroeconomic factor model underperforms a model that incorporates the residual market factor. For two of the three subperiods considered in the study, the mean errors from the macroeconomic factor model are significantly different from zero. The inclusion of the residual market factor results in a substantial reduction in mean errors and insignificance. Similarly to Van Rensburg (2000), Chang (1991) views larger mean residuals as reducing the power of statistical tests. The cross-sectional (APT relation) predictive ability of these specifications is also compared using Theil's U^2 statistic. The results again indicate that a specification that only incorporates macroeconomic factors has poor forecasting ability. Chang (1991) reports that the macroeconomic APT relation performs relatively poorly, even underperforming a naïve forecast. The inclusion of the residual market factor improves forecasting performance. The resultant U^2 statistic indicates that the specification combining the residual market factor and the macroeconomic factors outperforms the macroeconomic factor model and also a simple naïve forecast across all subperiods. Chang's (1991) findings suggest that factor omission will translate into inferior intertemporal and cross-sectional predictive ability (consequence 5) in Section 5.3.1.).

Clare *et al.* (1997b: 646, 648) investigate the consequences of departing from an assumed strict factor structure of uncorrelated residuals. This is done by applying the Fama and Macbeth two-step procedure which assumes a strict factor structure (Section 2.2.) and NL3SLS regression (Section 3.3.). The latter procedure permits the variance-covariance matrix of idiosyncratic returns to be estimated from a strict factor structure or an approximate factor structure that permits correlation of the residuals and therefore captures factors relegated to the residuals. When the Fama-MacBeth procedure is applied,[81] only two factors are priced, the Gilt to Equity Yield Ratio (GEYR) and the retail price index. When a strict factor structure is assumed and the NL3SLS methodology is applied, none of the factors

[81] Clare *et al.* (1997b: 649) use data from a prior study, that of Clare and Thomas (1994), and report the results of the Fama-MacBeth procedure from this study which considers a total of 18 macroeconomic factors.

considered are priced. Under an approximate factor structure, assumed under the NL3SLS procedure, five factors are priced. These are returns on the FTSE All Share Index, the retail price index, corporate default risk, the yield on an index of UK debentures and loans and a measure of retail bank lending. Clare *et al.* (1997b) attribute this to efficiency gains associated with not restricting the variance-covariance matrix to be diagonal and state that the specification of the covariance matrix in tests of the APT is an importance aspect. These results suggests that if a technique that assumes a strict factor structure is applied, such as the Fama-Macbeth approach, inferences based upon the resultant APT relation will be misleading.

The statistical APT precedes the macroeconomic APT. However, given its limitations, it is the macroeconomic APT that prevails in practice. A pertinent question that arises is whether a statistical or macroeconomic APT is more appropriate in explaining returns. Spyridis *et al.* (2012: 40) compare statistical and macroeconomic versions of the APT using data for the Athens Stock Exchange (ASE). A six factor structure is proposed based on principal component analysis. In cross-sectional tests, the $\bar{R}^2$ for the entire sample period is 39.9%.[82] The macroeconomic APT, which incorporates expected and unexpected inflation, the growth rate in industrial production, term structure, changes in petroleum prices, and returns on the ASE, has an $\bar{R}^2$ of 0.168. The authors attribute the poor performance of the macroeconomic APT to the changing nature of the ASE and suggest that new factors have emerged over time. The lower $\bar{R}^2$ for the macroeconomic APT over the entire period indicates that the statistical APT is more appropriate for explaining expected returns. Spyridis *et al.* (2012:55) confirm this by applying the Davidson and MacKinnon (1981) test. The reason cited for the dominance of the statistical APT is the presence of unobserved factors that potentially feature in asset pricing for this specific market. This also suggests that macroeconomic factors do not fully proxy for true factors that are statistically derived and unidentified.[83] Also, as indicated by the authors, this demonstrates a limitation of the macroeconomic APT; the relevance of factors changes over time resulting in potential underspecification.

[82] Tests are also conducted over subperiods and for different portfolios. Results are comparable with those for the entire period.

[83] Spyridis *et al.* (2012: 52) report that regressions of factor scores onto the macroeconomic factors indicate that there are significant relationships between the factor scores and the macroeconomic factors. However, as the statistical APT outperforms the macroeconomic APT, it seems that macroeconomic factors are inadequate proxies for derived factors (Middleton & Satchell, 2001: 506).

Early evidence that underspecification impacts the structure of residual variance is observed by Bera *et al.* (1988: 203), who apply an ARCH(1) specification to permit beta estimation in the market model to reflect heteroscedasticity (also see Engle, 2001: 160). The impact of conditional heteroscedasticity is investigated by estimating least squares and ARCH versions of the market model using CRSP return data and a value-weighted NYSE index. Results indicate that the betas in the ARCH versions of the model are more efficient and that coefficient estimates are impacted. Bera *et al.* (1988) report that the greater the level of conditional heteroscedasticity, that is heteroscedasticity dependent upon model specification, the greater the difference between the least squares and ARCH betas[84] in the market model. Moreover, the authors state that owing to the nature of the ARCH process, the impact of omitted factors is captured. The reflection of underspecification in the ARCH process can be seen as a matter of impure heteroscedasticity, namely heteroscedasticity that is attributable to omitted factors (Bucevska, 2011: 631). Koutoulas and Kryzanowski (1994: 342) investigate the integration of the Canadian and North American markets using national and international macroeconomic factors. Following the estimation of the linear factor model for returns on the TSE, equally- and value-weighted indices of stocks on the TSE and 50 size-ranked portfolios of Canadian stocks, the authors report that there is no evidence of the presence of ARCH errors. The authors state that this implies that "a well-specified return generating model captures a significant portion of time-varying volatility in stock returns" (Koutoulas and Kryzanowski, 1994: 342). Conversely, together with the argument of Bera *et al.* (1988), this implies that underspecification will impact the variance structure of the residuals and will induce heteroscedasticity. In the presence of heteroscedasticity, inferences drawn from a representation of the linear factor model estimated using least squares will be unreliable as standard errors will be underestimated (Baretto & Howland, 2006: 555-556, 560). It therefore follows that the impact of underspecification on the conditional variance structure should be considered and that the impact of factor omission can be investigated by analysing the conditional variance structure, as common factors may be reflected in residual volatilities (Renault, van der Heijden & Werker, 2016: 23).

Jorion (1991: 366) investigates the pricing of exchange rate risk in the US stock market using the APT framework. In formulating the specification, the author postulates that in underspecified APT models, the findings of significant risk premia can be interpreted in

[84] The ARCH betas are estimated using ML estimation.

terms of omitted factors. An example of inflation risk is cited, which may be a proxy for the exchange rate. While a given factor may appear to be priced (e.g. inflation), this relationship will no longer hold if other factors, for which this factor is a proxy, are included in the APT relation (e.g. exchange rate). Jorion (1991) departs from the use of a two-factor APT model, as used by Sweeney and Warga (1986) to study the impact of a specific factor (the interest rate in this specific study), and augments the model with the six Chen *et al.* (1986) factors against which the exchange rate (measured by innovations in a trade-weighted exchange rate) is orthogonalised. The author argues that = orthogonalisation avoids spurious pricing that arises from the exchange rate acting as a proxy for other underlying factors. The pricing of the (residual) exchange rate in cross-sectional industry returns is then investigated and exchange rate risk is found to be generally unrelated to expected returns. Jorion's (1991) approach suggests that underspecification of the linear factor model may translate into erroneous inferences relating to the pricing of factors in the APT relation. The coefficient estimated in the linear factor model may be priced because of correlation with omitted factors and not because of a specific factor's pricing ability.

Dominguez (1992: 87, 94) states that it may not be possible to identify and quantify all relevant factors. This will result in coefficient biases and a bias of the intercept. The author further investigates the sensitivity of the linear factor model to omitted factors. A model using daily industrial return data and three factors, namely returns on an equal-weighted NYSE index,[85] changes in the interest rate and the exchange rate is estimated.[86] Results indicate that the equally-weighted NYSE index is widely significant and that the exchange rate is significant for a substantial number of portfolios. In the next step, unanticipated changes in the exchange rate and the equally-weighted market index are found to be priced. The estimated intercept implies a risk-free rate of around 50%. Dominguez (1992) states that this is indicative of a bias attributable to factor omission. To investigate factor omission and to quantify the resultant bias, two potential (omitted) factors reported at a monthly frequency, namely the industrial production and the degree of risk aversion, are regressed onto the three factors included in the initial linear factor model specification.[87] Industrial

---

[85] The equal-weighted NYSE index is included to reduce the probability of misspecification, to nest the prediction of returns within the CAPM and to provide a test of the APT. If the CAPM is a correct description of returns, then no other factor should be significant, according to Dominguez (1992: 88).

[86] As daily data is used, the model is restricted to factors that are available on a daily basis and are of a more financial nature than macroeconomic nature. As a result of this, specification bias is introduced.

[87] Risk aversion is defined as the difference between low-grade bonds and long-term government securities (Dominguez, 1992: 96). This test assumes that if a factor is omitted, part of its influence will be accounted for

production and risk aversion are both significantly related to changes in the interest rate and risk aversion is also related to returns on the equally-weighted NYSE. Dominguez (1992: 91) argues that this points towards potential biases in the coefficients of the interest rate and the equally-weighted NYSE index in the original linear factor specification. However, as the exchange rate is not significant, the estimated risk premium (in the cross-sectional APT model) for this factor is unaffected, and the observed pricing relationships are not indicative of underlying omitted factors. Similarly to Jorion (1991), Dominguez' (1992) approach suggests that biases in the estimated coefficients have the potential to affect pricing relationships through their impact on the estimated risk premia.

Jorion (1991) and Dominguez (1992) emphasise the impact of factor omission on inferences relating to pricing. Panetta (2002: 444) states that inferences relating to the return generating process, as represented by the linear factor model, will be also impacted. In the investigation of the stability of the return generating process for the Italian stock market, Panetta (2002) provides yet another reason for the finding of instability between returns and macroeconomic factors (Section 2.4.), namely that the model of the return generating process may be misspecified and that this may be responsible for the observed instability. Results will be impacted if the omitted factors are not orthogonal to the factors that are included. Coefficient instability will be driven by the changing relationship between factors that are included and the omitted factor or factors. This instability will be reflected in changing coefficient bias for the factors that comprise the model but reflect the impact of correlated omitted factors.

Another consequence of underspecification is that the APT model can be rendered erroneously invalid. A central tenet of the APT is that only systematic factors are priced, as all idiosyncratic (analogously sector-specific) factors are diversifiable (Section 2.3.1.). This can be established by including an additional factor in the form of the standard deviation (or variance) of the residuals derived from the linear factor model in the APT model (Ross & Roll, 1980: 1093; Dominguez, 1992: 97).[88] If the APT is valid, expected returns will be explained by factor loadings on systematic factors and will not be related to idiosyncratic factors. Brennan *et al.* (1998: 349) argue that a finding of a priced idiosyncratic factors (or

by the other factors included in the model, if the omitted factor is indeed part of the true return generating process. The extent of the bias is determined by the strength of the correlation between the included and omitted factors.

[88] Other measures of idiosyncratic risk that have been used are size. See Chen (1983) and Yli-Olli and Virtanen (1992).

factors) may suggest that the APT is not invalid, but that an idiosyncratic factor is proxying for exposures to omitted factors that are priced. If the idiosyncratic factor used to test the validity of the APT model is the residual variance or standard deviation, then this is particularly relevant. If the linear factor model is underspecified, then the residual variance will reflect dispersion that is attributable to omitted systematic factors and not purely idiosyncratic factors. Residual variance will therefore be associated with a significant risk premium because systematic factors are relevant in the pricing relation, as opposed to idiosyncratic factors, and this may result in an erroneous rejection of the validity of the APT model (Lehmann, 1990: 72; Dominguez, 1992: 98). Elton *et al.* (1995: 1239) provide an apt summation relevant to this discussion. The authors suggest that the APT model will fail if the linear factor model is underspecified or if it truly does not hold. Such tests of the validity of the APT relation are also joint tests of the return generating process. Underspecification will confound pronouncements on the validity of the APT with those on the adequate specification of the return generating process.

Middleton and Satchell (2001: 503-506) argue that given that the number of factors in the linear factor model is uncertain, practitioners are prone to make unavoidable errors about the identity and the number of factors. Through derivation, the authors show that APT pricing errors will, on average, be zero if the true number of factors can be represented by an equivalent or greater number of proxy factors that are correlated with the true factors. However, if the number of proxies is less than the true number of factors, the average pricing error may not be zero although an APT relationship may be derived. In outlining implications, Middleton and Satchell (2001) state that the problem of underspecification in the APT framework can only be avoided if factors are derived statistically and a sufficiently significant number of factors is arrived at. If macroeconomic factors are used in place of statistical factors as proxies, the problem of underspecification persists. To arrive at a correct APT pricing relation (one with zero average pricing errors), the number of correlated macroeconomic proxy factors must be either equal to or greater than the number of factors derived from returns. Consequently, the use of a generous rather than parsimonious number of factors is recommended when constructing linear factor models. The difficulty in identifying highly correlated macroeconomic proxy factors, deliberate parsimony, data unavailability and the unobservability of factors poses a challenge to including the required number of factors in the linear factor model (Section 5.2.). In conclusion, Middleton and Satchell (2001), through derivation, suggest that underspecification will introduce APT pricing errors and emphasise the difficulties associated with specifying the true return generating process.

Similarly to Middleton and Satchell (2001), Van Rensburg (2000: 36-37) states that if macroeconomic factors are used as explanatory factors, the linear factor model is likely to be underspecified and that the assumption of uncorrelated residuals across assets is likely to be violated (Section 3.4.). Importantly, the result will be such that the estimators of variance will be upwardly biased, resulting in potentially erroneous failures to reject the null hypothesis of no relationship between returns and a factor(s). It is shown that underspecification can lead to misidentification of relevant factors in the linear factor model. In a preliminary screening of candidate factors, returns on the JSE All Share Index are found to be significantly correlated with five factors, namely the JSE All Share Index Earnings, the Dow Jones Industrial Index, the yield on the 10 year government bond, the level of gold and foreign exchange reserves and the Rand gold price. A regression of returns on the JSE All Share Index onto these five factors indicates only three significant relationships. When the model is re-estimated with the two residual market factors included (derived from the JSE All-Gold and Industrial indices), the standard errors associated with the coefficients decrease (by more than half) for all factors and all five factors are now significant. Expectedly, the explanatory power, as measured by the $\bar{R}^2$, triples (from 0.29 to 0.91). Van Rensburg (2000: 37) states that had the initial specification not been augmented with these two factors, the "resulting estimation bias would mislead the researcher into inferring that these variables do not contribute to explaining the time series variation in equity returns."[89]

In a subsequent study, Van Rensburg (2002: 91-92) states that prior studies have ignored the violation of the residual diagonality assumption. It is further argued that models that ignore this assumption are characterised by specification biases. The author states that it can also be shown that the variance of the intercept term is similarly biased upwards. The consequence in a more practical setting is that if excess returns are used, a model may suggest that that there is no under or over-performance for an asset or a portfolio when it does exist (also see Lehmann & Modest, 1987). By using a single factor (the money market shortage), returns on a South African conglomerate (the Rembrandt Group) and two residual market factors (the residualised Financial-Industrial and Resources indices), it is again

[89] As the *t*-statistics used in significance tests are calculated by dividing the estimated coefficients by the standard errors and the standard error is inflated as a result of underspecification, the result is a potentially erroneous failure to reject the null hypothesis.

demonstrated that underspecification leads to the misidentification of a significant relationship. The explanatory power of the model increases dramatically, as before. The broader impact of this misidentification is well-articulated by Ferson and Harvey (1994: 785). The authors, in their study of the sources of risk in global equity returns, state that if a coefficient for a given factor is not significantly different from zero, then it is omitted from analysis and that as a result, there can be no risk premium associated with this factor in the APT model. This is because such a factor will not be included in the linear factor model in the first place. Most importantly, Van Rensburg (2002) recognises that there is a lack of awareness of the consequences of violating the diagonality assumption as a result of underspecification.

5.4.3. The Importance Of Investigating Underspecification

The literature in this section indicates that underspecification impacts various aspects of the linear factor model and the APT. Underspecification can result in an erroneous failure to reject a null hypothesis of no relationship between returns and macroeconomic factors in the linear factor model, incorrect inference making, poor predictive performance, heteroscedasticity in the residuals and a general misidentification of the return generating process. Most concerningly, underspecification can lead to the omission of relevant factors from the APT relation, incorrect pronouncements on the pricing of specific factors and even an erroneous rejection of the validity of the APT as a theory.

The building block of the APT is the linear factor model. Therefore, the problem of underspecification begins with the linear factor model and extends into the APT relation (Elton & Gruber, 1997: 1750). As suggested by Van Rensburg (2002), there is a lack of awareness about the consequences of underspecification of the linear factor model.

Middleton and Satchell (2001: 506) state that the problem of underspecification is likely to persists if macroeconomic factors are used in the linear factor model. APT literature widely relies upon the residual market factor to proxy for omitted factors and to resolve underspecification (Section 3.2.; Section 3.4.). Nevertheless, there is a gap in the literature; the adequacy of the macroeconomic linear factor model in describing the return generating process and the consequences of underspecification on the linear factor model are not investigated comprehensively and jointly. The ability of the residual market factor to resolve underspecification is also not comprehensively considered in the literature. Given the foundational role of the macroeconomic linear factor model in APT literature and the widespread use of the residual market factor to resolve underspecification, the consequences of underspecification should be considered further. This can be done by investigating the consequences of underspecification and then establishing whether the inclusion of a residual market factor or two residual market factors alleviates the associated symptoms.

5.5. CHAPTER SUMMARY AND CONCLUSION

This chapter begins by listing possible causes of underspecification. There are a number of reasons why relevant factors may be omitted from a linear factor model specification, intentionally or unintentionally (Section 5.2.). Underspecification results in an inconsistent and biased intercept and coefficient estimates, inflated residual variance, misleading hypothesis tests, adversely impacts model predictions and can induce heteroscedasticity and serial correlation into the residuals (Section 5.3.1.). The immediate consequences of underspecification are illustrated in Section 5.3.2.This example indicates that underspecification can impact the general outcome of an empirical investigation.

Section 5.4.1. briefly discusses the impact of underspecification on the assumptions underpinning the APT linear factor model. Factor omission results in the violation of the diagonality assumption and endogeneity. Sections 5.4.2. demonstrates that the consequences of underspecification are evident in the linear factor model and the application of the APT. Notably, factor omission results in a misidentification of relevant factors in the linear factor model, incorrect inferences relating to priced factors and even a potentially erroneous rejection of the APT model. The use of macroeconomic factors is unlikely to yield an adequate description of the linear factor model and therefore, a residual market factor is widely used in the literature to resolve underspecification in macroeconomic models motivated by the APT framework. As the linear factor model is a building block of the APT, the consequences of underspecification and the ability of the residual market factor to resolve underspecification should be considered further (Section 5.4.3.).

The adequacy of the residual market factor (or factors) in resolving underspecification can be investigated by establishing whether the inclusion of the residual market factor mitigates the consequences of underspecification. This requires an understanding of the consequences of underspecification, which are discussed in this chapter. Chapter 6 sets out the data and methodology used in investigating underspecification in the linear factor model and the ability of the residual market factor to account for omitted factors and to resolve factor omission.

CHAPTER 6

DATA AND METHODOLOGY

6.1. INTRODUCTION

This chapter outlines the data and the methodology applied in investigating underspecification in the macroeconomic APT linear factor model and the ability of the residual market factor to proxy for omitted factors and to resolve underspecification. The concepts discussed in the preceding chapters feature prominently in this chapter and the chapters that follow.

Studies that employ pre-specified macroeconomic factors to represent pervasive influences in returns rely upon the residual market factor to fulfil the role of a catch-all proxy for omitted factors (Section 2.3.; Section 2.4.). However, the literature indicates that a single residual market factor may be insufficient to account for omitted factors (Section 3.4.). A candidate for a second residual market factor is the MSCI World Market Index. A review of the literature in Chapter 4 indicates that markets are interdependent and macroeconomic news may impact multiple markets, which are external to the market from which the news originates. This suggests that a proxy for global influences should be considered in the linear factor model. Moreover, the presence of a second significant proxy in the linear factor model, aside from the conventional residual market factor, implies that macroeconomic factors and a single residual market factor are unable to adequately proxy for all pervasive influences in returns. Therefore, the inclusion of a second residual market factor constitutes a test of the adequacy of the conventional residual market factor. Chapter 5 outlines and emphasises the direct and the more subtle consequences of underspecification and the implications for inference and pricing within the APT framework.

The approach followed in this study is comparative (Section 1.3.). A benchmark model, which comprises macroeconomic factors, two residual market factors and a factor analytic augmentation is estimated. Three more specifications are then estimated. The first is a restricted specification comprising only macroeconomic factors. The next two are the unrestricted models, which incorporate the conventional residual market factor in addition to the macroeconomic factors and subsequently, a second residual market factor derived from the MSCI World Market Index. These are referred to as the unrestricted market model and the unrestricted model respectively, and collectively as the unrestricted models. Comparisons are made across specifications with the intention of summarising the consequences of underspecification and establishing whether the inclusion of the residual market factor and the second residual market factor mitigates the consequences of factor omission.

This chapter proceeds to set out the data that will be utilised in this study in Section 6.2. Also reported in this section is a preliminary analysis of the data. The methodology applied in investigating the factor structure of the data is set out in Section 6.3. The factor structure is investigated to determine the number of pervasive factors in returns and to assist in the identification of macroeconomic factors that are proxies for the pervasive influences in stock returns. Section 6.4. outlines the specifications considered, the econometric methodology employed and the various aspects that are considered in establishing the impact of factor omission and the ability of the residual market factors to resolve underspecification. Section 6.5. summarises the chapter and concludes.

6.2. DATA

6.2.1. Return And Macroeconomic Data

Financial data in the form of industrial index levels is obtained from the IRESS Expert database which provides financial and other data for South Africa and other African countries. The financial data is for industrial sectors that comprise the South African stock market, the JSE. As of 2017, the South African stock market features in the top 20 stock markets in the world by market capitalisation. It has a market capitalisation of $951 billion, is the largest stock market on the African continent and has been operational for 128 years with 472 current listings. In terms of market capitalisation, it exceeds the Madrid Stock Exchange, the Taiwan Stock Exchange and the BM&F Bovespa (Brazil) (Desjardins, 2017). The use of industrial sector data is advantageous in that it provides wide market coverage, does not need adjustments for corporate actions, avoids survivorship bias and illiquidity and permits for it to be established whether factors are pseudofactors, i.e. important for certain sectors and therefore not systematic by definition (Hughes, 1984: 42). Furthermore, Cho (1984: 1492) states that the use of industrial sector data ensures that common factors are extracted as the likelihood of extracting pseudofactors is minimised because of the distinctiveness of industrial sectors.

The sample period spans the period January 2001 to December 2016, yielding a total of 192 months of data. Only industrial sectors with a full data history are included in the sample and this constitutes 26 industrial sectors (out of a total of 33 sectors at the time of writing) comprising the sample. Month-end data is used and the risk free rate used to derive excess returns is the closing yield on the R186 government bond.[90] The length of the sample is motivated by the desire to capture a mix of long-term and short-term economic trends and dynamics and a more practical consideration in the form of the availability of macroeconomic data in South Africa since 2000.

Table 6.1. presents the industrial sectors included in the sample, the economic sectors to which these industrial sectors belong and the corresponding JSE index codes.

Table 6.1: List Of Industrial Sectors

Economic Sector	Industrial Sector	Index Code
Basic Materials	Chemicals	J135
	Forestry & Paper	J173
	Ind. Metals & Mining	J175
	Mining	J177
Industrials	Constr. & Materials	J235
	General Industrials	J272
	Elec. & Elec. Equip.	J273
	Indust. Engineering	J275
	Indust. Transp.	J277
	Support Services	J279
Consumer Goods	Automobiles & Parts	J335
	Beverages	J353
	Food Producers	J357
Health Care	Health Care Equip. & Services	J453
	Pharm & Biotech.	J457
Consumer Services	Food & Drug Retailers	J533
	General Retailers	J537
	Media	J555
	Travel & Leisure	J575
Telecommunication	Fixed Line Telecoms.	J653
Financials	Banks	J835
	Non-life Insurance	J853
	Life Insurance	J857
	General Financial	J877
	Equity Investment Instruments	J898
Technology	Software & Comp. Serv.	J953

Continuously compounded total monthly returns are used, defined as the natural logarithm of industrial sector returns (Tsay, 2002: 4):

$$r_{it} = \ln S_{it} - \ln S_{it-1} \tag{6.1}$$

[90] This specific proxy for the risk-free rate finds support in both academia and in practice (Nel, 2011: 5342; PWC, 2015; 44).

where r_{it} is the total return on industrial sector index *i* at time *t* and S_{it} is the level of index *i* at time *t*. Excess total returns, R_{it}, are obtained by subtracting the risk-free rate (the R186) from the logarithm of total returns in equation (6.1).

Figure 6.1. and Figure 6.2. depict the JSE All Share Index levels and returns over the sample period:

Figure 6.1: JSE All Share Index Levels: January 2001 to December 2016

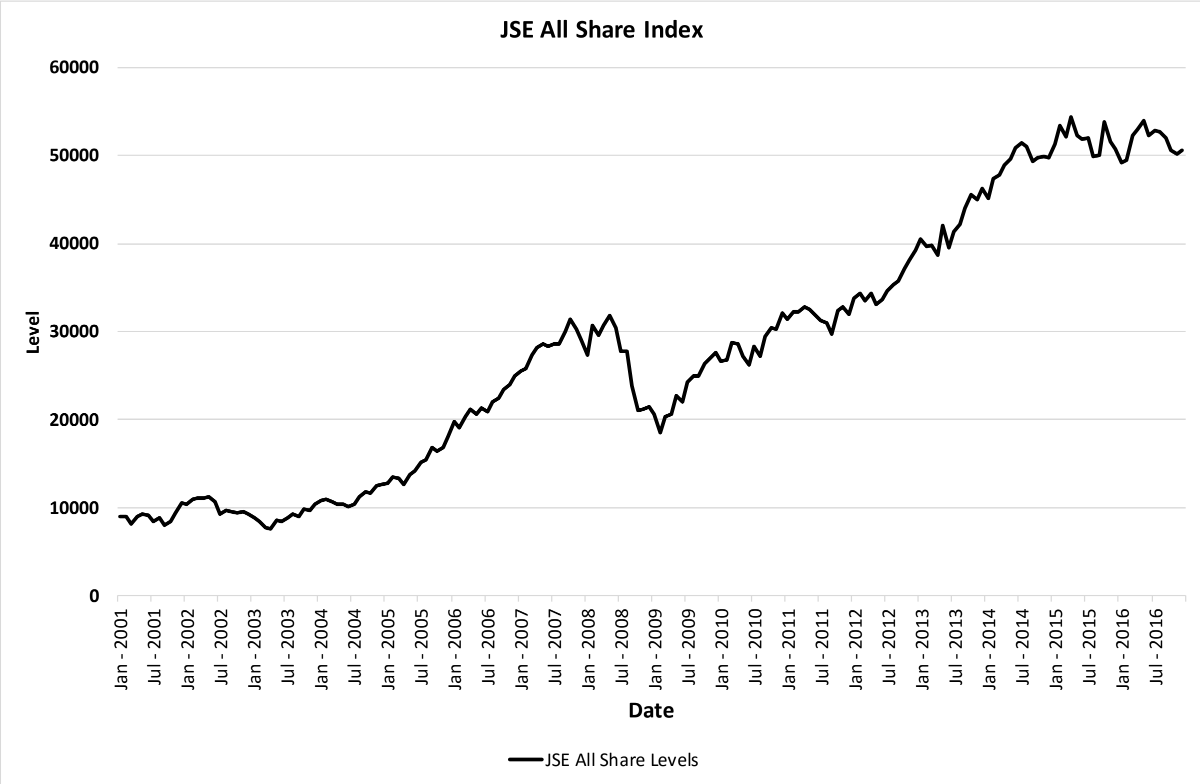
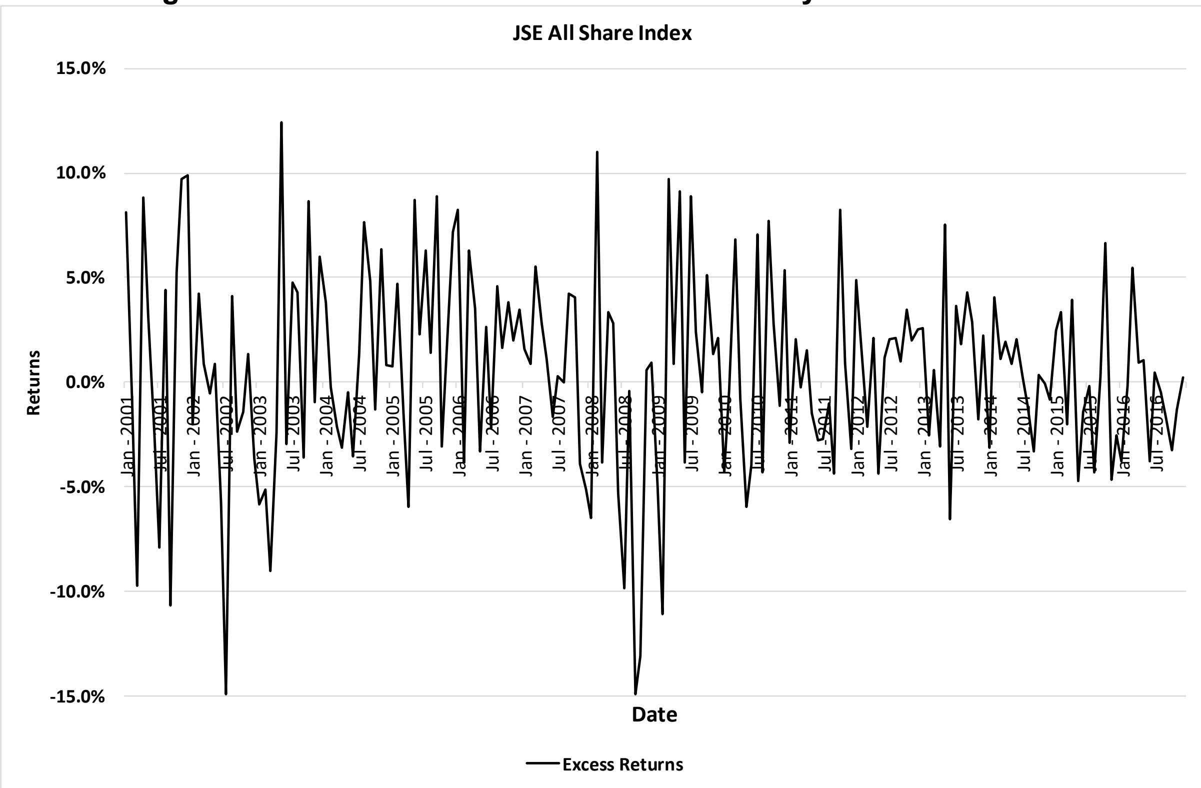


Figure 6.2: JSE All Share Index Returns: January 2001 to December 2016



The sample period coincides with numerous notable economic events that have a potential impact on the South African stock market, given its integration with global financial systems. These include the terrorist attack on the World Trade Centre in 2001, the war on terror in Iraq (2003), the ongoing emergence of China and India as economic powers, the Russian financial crisis (2008-2009, 2014-present), the sub-prime mortgage crisis of 2007 and 2008 and the Great Recession that follows, the European sovereign debt crisis (2009 onwards) and the early stages of the British withdrawal from the European Union (2016 onwards) (Peavler, 2018). South Africa has also experienced changing socio-economic conditions during this period. These are a commodities driven economic boom prior to 2008, increasing trade with India and China (Broadman, 2008: 96-97), a recession for the first time since 1994, the hosting of the FIFA World Cup in 2010 and a shifting political landscape (2016 municipal elections).

The Quantec EasyData database is used to obtain monthly macroeconomic data to construct the macroeconomic factors that are hypothesised to represent pervasive factors in stock returns (Chen *et al.,* 1986: 384-385; Connor, 1995). For ease of reference and for the purposes of model construction, which is described later, each factor is classified under an expanded classification similarly as in Elton and Gruber (1988: 41) and Hanousek and

Kočenda (2011:177 – 180) (see Table 6.4. in Section 6.2.3. for a comprehensive list of factors considered):

1) Real Activity
2) Prices
3) Cyclical Indicators
4) Exchange Rates
5) Monetary Factors
6) Commodities
7) Interest Rates
8) Trade
9) Market Indices

Although the APT does not directly provide guidance as to which factors should enter the linear factor model, the literature widely refers to the dividend discount model to identify factors that are assumed to proxy for the systematic factors that drive returns in the APT framework (Azeez & Yonezawa, 2006: 577; Section 2.3.4.). The dividend discount model, stated here for comprehensiveness, is used in a preliminary screening of potential factors and factors that can reasonably be seen as impacting expected cash flows or the discount rate or both are considered:

$$S_{it} = \sum_{t=1}^{\infty} \frac{E(D_{it})}{(1+r)^t} \tag{6.2}$$

where S_{it} is the price level for asset *i* at time *t*, *E* is an expectation operator relating to expectations about future cash flows as denoted by D_{it}, and the discount rate prevailing between *t* and $t+\tau$ is denoted by *r*. The drawback of this approach is that although it guides the identification of factors, some factors may not have a pervasive impact as required by the APT and therefore are not true APT factors. Such factors, although macroeconomic in nature, may be trivial (with low explanatory power) or may be non-trivial but not pervasive in that they only impact certain industrial sectors (Kryzanowski & To, 1983: 42). Therefore, the reliance on the dividend discount model to screen factors should be seen as a preliminary step in the identification of a broad but yet unconfirmed set of macroeconomic factors.

The APT does not explicitly prescribe how factors should enter the linear factor model. It can be readily inferred that the APT requires factors to be stationary time series, given that the dependent factor in the linear factor model, namely returns, is of a stationary nature. Three differencing techniques are used in this study to derive stationary times series from macroeconomic factors that are integrated of order one (*I(1)*). The first technique involves taking the first differences of logarithms ($f_{kt} = \ln L_{kt} - \ln L_{kt-1}$) (DL) in factor levels (L_{kt}). The second involves taking the first differences of the factor levels ($f_{kt} = L_{kt} - L_{kt-1}$) (D) and the final technique uses percentage changes in factor levels ($f_{kt} = (L_{kt} - L_{kt-1}) / L_{kt-1}$) (PC) (Clare & Thomas, 1994: 313).

The assumption underlying the APT is that stock prices react to macroeconomic announcements implying that news is unanticipated. This requires that macroeconomic factors should enter the linear factor model as innovations/unanticipated changes (Azeez & Yonezawa, 2006: 579). Therefore, a factor must be completely unpredictable at the beginning of each period, implying that a factor cannot be forecast on the basis of prior information or other publicly available information. Such a factor will have an expected value of zero, $E(f_{kt}) = 0$ (Berry *et al.,* 1988: 30). It can be established from an examination of the correllograms for a factor series that a specific factor is stated in innovations which represent unpredictable components. If factors are predictable and therefore are not white noise innovations, unexpected components that are zero-mean innovations must be generated with an appropriate methodology (Priestley, 1996: 872; Clare & Priestley, 1998: 107). Aside from deviating from a key requirement of the APT framework, failure to use true[91] innovations may give rise to spurious relationships and an errors-in-variable problem (Poon & Taylor, 1991: 624). To derive innovations, correllograms are inspected and if significant correlation is detected at a specific order(s), up to 12 orders, an autoregressive time series model is used to remove components that permit predictability by incorporating these lags into an autoregressive model. The residuals of this specification are now a representation of the unexpected components in the factor series (Priestley, 1996: 875; Van Rensburg, 1996: 106). Under the autoregressive time series methodology, the hypothesised expectations generating process can be represented as follows:

[91] Priestly (1996: 875) shows that a popular methodology, the rate of change methodology which involves the simple differencing of macroeconomic factors, fails to generate innovations and that the presumed unexpected components generated using this methodology are serially correlated. Chan *et al.* (1985) and Chen *et al.* (1986) apply this methodology and therefore do not use true innovations.

$$f_{kt} = \alpha + \sum_{k=1}^{K} b_k f_{kt-\tau} + \varepsilon_{kt} \tag{6.3}$$

where f_{kt} is the observed level of macroeconomic factor *k* at time *t* and $\sum_{k=1}^{K} b_k f_{kt-\tau}$ is the set of lagged factor terms identified as predicting the level of factor *k* at time *t,* with the associated betas. The residuals, ε_{kt}, of equation (6.3) are used in place of the original differenced series to represent innovations in the respective macroeconomic factors. The macroeconomic factor series correspond to the sample period length set out above (January 2001 to December 2016) but include an additional 12 months of data prior to the start of the sample on January 2001. Innovations are therefore derived using a sample that includes an additional 12 months of data, from January 2000 to December 2000, to ensure that the construction of innovations does not consume degrees of freedom in the actual sample used in the analysis (Van Rensburg, 2000; 33).

A preliminary analysis of the return and macroeconomic data is reported in Section 6.2.3. with the aim of gaining insight into the statistical properties and readiness of the data for use in model construction. The next section sets out the methodology employed for the purposes of statistical analysis.

6.2.2. Methodology Used In The Preliminary Analysis Of The Data

It is well documented that financial time series are not “well-behaved.” Return series deviate from the classical assumptions of normally, independently and identically distributed (*n.i.i.d*) returns. Xiao and Aydemir (2007: 3) state that distributions of financial time series exhibit fatter tails than those of a normal distribution and that the kurtosis for many financial time series is above three (returns are leptokurtic). Akgiray (1989: 60) rejects the assumption of independence for returns on the CRSP value-weighted index. Taylor (2008:55) shows that a selection of US stocks exhibits non-constant variance. Engle (2001: 158-159) suggests that returns on the Dow Jones and NASDAQ indices exhibit volatility clustering. Roll (1992: 31) states that heteroscedasticity and leptokurtosis can lead to incorrect inferences as estimated standard errors will be biased and inconsistent. This emphasises the need to consider the distributional properties of the return data. Furthermore, for macroeconomic factor series, Priestley (1996: 875) shows that a simple method of obtaining changes in factor levels, namely the rate of change methodology, fails to the produce the innovations that are required by the APT. This emphasises the need to outline the statistical characteristics of the generated factor series. Different aspects of the data are important for the return and macroeconomic factor innovation series and therefore different aspects are reported in Section 6.2.3. The overview of the methodology that follows highlights the aspects considered.

The Jarque and Bera (JB) (1980, 1987) test is applied to determine whether a given series conforms to the normality assumption. This test establishes whether the distribution is jointly characterised by a kurtosis (*K*) coefficient of three and a skewness (*S*) coefficient of zero (Cryer & Chan, 2008: 282). The JB test statistic is defined as follows (Thadewald & Büning, 2007: 91):

$$JB = \frac{n}{6}\left(S^2 + \frac{(K-3)^2}{4} \right) \tag{6.4}$$

where *n* is sample size, *S* is sample skewness and *K* is sample kurtosis. Departures from normality in the return series are of interest; the return series are the dependent factor in the analysis and departures from normality have the potential to impact the results of the analysis. For example, it is recognised in the literature that the presence of leptokurtosis is indicative of non-stationary variance and findings of departures from normality guide the use of the appropriate econometric methodology (Akgiray, 1989: 62). Box plots are used to identify outliers in returns and far (extreme) outliers are excluded in the testing of the normality assumption as these may bias the test in favour of a rejection of the null hypothesis of normality. Near outliers are retained as these may be the result of volatility clustering and not unusual events (Hodge & Austin, 2004: 9).

The validity of the assumption of statistical independence for returns and the presence of dependence in the macroeconomic factor series are both investigated. The investigation into the independence (or the lack of) of the return series begins with the serial correlation model (Fama, 1965: 68-69; Campbell, Lo & MacKinlay, 1997: 44):

$$\rho_\tau = \frac{\mathrm{cov}(R_{it}, R_{it-\tau})}{\mathrm{var}(R_{it})} \tag{6.5}$$

where ρ_τ is the serial correlation coefficient, τ is the lag order and R_{it} is the return on series *i* at time *t*. Statistically significant correlation coefficients are indicative of whether the assumption of independence holds and the magnitude of dependence at individual lag orders, if present (Fama, 1965: 81). Further testing, for both the return series and macroeconomic factor series, is conducted using Ljung-Box *Q*-statistics which indicate whether serial coefficients up to a certain order are jointly equal to zero (Ljung & Box, 1978):

$$LB = n(n+2)\sum_{\tau=1}^{m}\left(\frac{\rho_{\tau}^{2}}{n-\tau}\right) \tag{6.6}$$

where n is the sample size, m is the number of lag orders considered and ρ_{τ} is the serial correlation coefficient at lag order τ (now derived for the return series and the macroeconomic factor innovation series). For returns, this test is applied at five and 10 orders of serial correlation (Campbell *et al.,* 1997: 66). It is anticipated that for the return series, the assumption of independence will generally hold although isolated deviations may occur (Kendall & Hill, 1955: 18; Fama, 1965: 74). Poon and Taylor (1991: 625) state that significant serial correlation may be suggestive of a non-synchronous trading effect, thin trading or the presence of a common factor. It is hoped that the use of industrial indices will mitigate the two former issues.

For the innovations derived from the macroeconomic factor series, this test is applied for 12 orders of serial correlation. The assumption of independence should hold as components that permit predictability are reflected in the autoregressive model used to generate innovations. To confirm that the pre-whitening process (equation (6.3)) generates true innovations, the Breusch-Godfrey serial correlation LM test is also applied to test for serial correlation up to the 12th order (Godfrey, 1978; Breusch, 1978; Clare & Thomas, 1994: 315). In this test, the residuals of equation (6.3) are regressed onto 12 lags:

$$f_{kt} = \alpha_0 + \sum_{k=1}^{K}\rho_k f_{kt-\tau} + \upsilon_{kt} \tag{6.7}$$

where f_{kt} is now the residual series derived from equation (6.3) ($f_{kt} = \varepsilon_{kt}$), representative of innovations in factor *k.* The null hypothesis is:

$$H_0 : \rho_1 = \rho_2 = \rho_3 = ... = \rho_k \tag{6.8}$$

and the alternative hypothesis is:

$$H_1 : \rho_1 \neq \rho_2 \neq \rho_3 \neq ... \neq \rho_k \tag{6.9}$$

where ρ_k is the coefficient on innovations in equation (6.7), associated with lag order τ. The null hypothesis implies that innovations are unrelated to previous residual terms. In other words, the pre-whitened observations are pure innovations (Brooks, 2008: 148).

Return series and the explanatory macroeconomic factors enter the linear factor in stationary form (Clare & Thomas, 1994: 316). The Augmented Dickey-Fuller (ADF) unit root test is applied to confirm the stationarity of each return and macroeconomic factor innovation series (Dickey & Fuller, 1979):

$$y_t = \delta y_{t-1} + \sum_{i}^{m} \alpha y_{t-\tau} + \varepsilon_t \tag{6.10}$$

where y_t represents a factor series (either the return series for industrial sector *i* or a series of innovations in factor *k*), t is the time trend and τ is the lag order. The ADF test incorporates a sufficient number of terms of factor y_t to ensure that the residual terms, ε_t, are uncorrelated. The null hypothesis for the ADF test is that δ is equal to zero. A failure to reject the null hypothesis is indicative of a unit root and that the series is non-stationary. An additional confirmatory non-parametric test that is applied is that of Phillips and Perron (1988) (the Phillips-Perron (PP) test) (Sadorsky & Henriques, 2001: 203). A failure to account for non-stationarity in the series may lead to spurious regression results (Van Rensburg, 1999: 36; Gujarati & Porter, 2009: 384). Near outliers in returns are retained so as not to extensively deviate from the true structure of the data and extreme outliers are not excluded in these tests of stationarity and in the exploration of the serial correlation structure of the data.

For a series to be stationary, both the mean and the variance must be stationary over time. To investigate the stationarity of the return variance, the *Q*-statistic is reported for squared return series, a proxy for volatility. A significant *Q*-statistic indicates ARCH effects in return data (Cryer & Chan, 2008: 282). *Q*-statistics are reported for the first 15 serial correlation coefficients for squared returns. The ARCH Lagrange Multiplier (LM) test is also applied to test for ARCH effects in returns. To test for ARCH effects in the return series in this study, squared residuals from a least squares regression are regressed onto a constant and prior lagged squared residual terms for the null hypothesis (Engle, 1982: 999):

$$H_0 : \alpha_{i1} = \alpha_{i2} = \alpha_{i3} = \ldots = \alpha_{ip} \tag{6.11}$$

where the alternative hypothesis is:

$$H_1 : \alpha_{i1} \neq \alpha_{i2} \neq \alpha_{i3} \neq ... \neq \alpha_{ip} \tag{6.12}$$

where the α s are the coefficients on lagged squared residual terms for sector *i* and *p* is the ARCH effect order. A rejection of the null hypothesis in equation (6.11) implies that there are ARCH effects in the residuals and that the variance of the residuals differs across time. Therefore, significant serial correlation in the squared residuals and the presence of ARCH effects are indicative of time-varying variance, volatility clustering and conditional heteroscedasticity. To generate the residuals for the ARCH LM test, an autoregressive model with a single lag (AR(1)) is estimated to represent a simple return generating process (Akgiray, 1989: 65):

$$R_{it} = \alpha_0 + b_1 R_{it-1} + \varepsilon_{it} \tag{6.13}$$

where R_{it} is the return on industrial index *i* at time *t* and R_{it-1} is the autoregressive term. The ARCH LM test is applied to test for ARCH(1), ARCH (5) and ARCH(10) effects in the residuals of equation (6.13). In the presence of ARCH effects, the least squares assumption of homoscedasticity of the residual terms will be violated. This will translate into a false sense of precision and inefficient coefficient estimates (Elyasiani & Mansur, 1998: 548; Wong & Bian, 2000: 65; Sadorsky & Henriques, 2001: 203). If preliminary results indicate that returns exhibit time-varying variance and this carries over into the residuals, as indicated by the ARCH LM test, then what Engle (2001: 157) terms “the workhorse of applied econometrics,” namely the least squares model, may not be appropriate for estimating the linear factor model.

6.2.3. Statistical Properties Of Return And Macroeconomic Data

Table 6.2. indicates that most return series exhibit some level of leptokurtosis. With the exception of the media sector, all series are characterised by a kurtosis coefficient that exceeds three. Most return series are negatively skewed with the exception of the industrial metals and mining, industrial engineering, food producers and non-life insurance sectors which exhibit positive skewness. On the basis of the JB test, the null hypothesis of normally distributed returns is rejected for 17 of the 26 industrial sector series and the JSE as a whole as measured by returns on the JSE All Share Index. As expected, the results of the ADF and PP tests indicate that all series are stationary.

The results in Table 6.3. indicate that the assumption of independence holds for the JSE All Share Index. The first five serial correlation coefficients are individually statistically insignificant and *Q*-statistics indicate that the first five and 10 serial correlation coefficients are not jointly significant. For industrial sector returns, independence appears to be a working albeit imperfect approximation of the serial correlation structure. The null hypothesis that the first five and 10 serial correlation coefficients are jointly equal to zero is rejected for nine and eight of the 26 sectors respectively. However, an analysis of the individual correlation coefficients indicates that joint significance may be driven by isolated instances of significant individual serial correlation. For example, only the second and sixth (not reported in Table 6.3.) serial correlation coefficients are significant for the non-life insurance sector yet both *Q*-statistics are statistically significant. Moreover, the magnitude of these coefficients is relatively low; ρ_2 is -0.161 and ρ_6 is 0.164.

The *Q*-statistic for the first 15 serial correlation coefficients of the squared return series (a proxy for volatility) in Table 6.3. is statistically significant for the JSE All Share Index and 12 of the industrial sector series suggesting that volatility is of a time-varying nature and exhibits volatility clustering (Engle, 2001: 162; Cryer & Chan, 2008: 278). The presence of higher order ARCH effects, ARCH(5) and ARCH(10) effects, is observed for the JSE All Share Index and 15 and 12 industrial sectors respectively. This suggests that the non-stationarity of the variance carries over into the residuals and that residuals exhibit evidence of autoregressive conditional heteroscedasticity.

Table 6.2: Statistical Properties Of South African Stock Returns

	Obs.	Mean.	Std Dev.	Kurt.	Skew.	JB Stat.	ADF	PP
JSE All Share Index	192	0.005	0.048	3.631	-0.305	6.156**	-13.991***	-14.056***
Chemicals	191	0.007	0.053	4.237	-0.503	20.230***	-14.231***	-14.229***
Forestry & Paper	191	0.006	0.083	3.018	-0.327	3.410	-13.884***	-13.925***
Ind. Metals & Mining	190	0.007	0.110	4.276	0.220	14.422***	-7.501***	-13.718***
Mining	192	0.001	0.081	3.506	-0.225	3.700	-14.269***	-14.297***
Const. & Materials	192	0.002	0.070	3.876	-0.495	13.994***	-11.528***	-11.511***
General Industrials	192	0.009	0.051	4.071	-0.514	17.637***	-13.802***	-13.808***
Elec. & Elec. Equip.	192	0.001	0.055	4.297	-0.460	20.230	-11.889***	-11.833***
Indust. Engineering	190	0.013	0.060	4.547	0.144	19.615***	-11.621***	-11.747***
Indust. Transp.	192	0.004	0.065	4.245	-0.591	23.581***	-13.741***	-13.762***
Support Services	192	0.000	0.056	3.370	-0.468	8.087**	-13.050***	-13.046***
Automobiles & Parts	191	0.001	0.085	3.887	-0.095	6.548**	-12.292***	-12.279***
Beverages	192	0.008	0.060	4.132	-0.117	10.692***	-14.313***	-14.300***
Food Producers	192	0.009	0.044	3.115	0.090	1.655	-12.881***	-12.883***
Health Care Equip & Services	192	0.013	0.059	3.939	-0.128	7.576**	-12.609***	-13.095***
Pharm & Biotech.	192	0.013	0.073	3.486	0.068	2.036	-13.638***	-13.686***
Food & Drug Retailers	192	0.010	0.058	3.622	-0.218	4.608***	-14.396***	-14.401***
General Retailers	192	0.008	0.066	3.008	-0.220	1.554	-12.273***	-12.246***
Media	191	0.017	0.081	2.941	-0.293	2.757	-13.223***	-13.266***
Travel & Leisure	192	0.005	0.055	4.173	-0.629	23.679***	-12.365***	-12.456***
Fixed. Line Telecoms	192	0.003	0.092	3.545	-0.416	7.899**	-12.579***	-15.597***
Banks	192	0.005	0.062	3.277	-0.061	0.735	-14.780***	-14.828***
Non-life Insurance	191	0.006	0.055	3.149	0.196	1.394	-13.453***	-13.483***
Life Insurance	192	0.002	0.059	4.140	-0.483	17.859***	-13.660***	-13.665***
General Financial	192	0.003	0.064	4.402	-0.192	16.897***	-12.106***	-12.079***
Equity Invest. Instruments	192	0.005	0.044	4.023	-0.195	9.584***	-15.312***	-15.241***
Software & Comp. Serv.	189	0.002	0.088	4.817	-0.428	31.773***	-11.407***	-11.453***

Notes:

The asterisks, ***, ** and *, indicate statistical significance at the respective 1%, 5% and 10% levels of significance. Obs. refers to the number of observations. Numbers below 192 indicate that extreme outliers have been omitted. Extreme outliers are identified using box plots and are omitted for distributional tests. Mean and Std Dev. are the respective mean and standard deviation values for each return series. Kurt. and Skew. are the respective kurtosis and skewness coefficients for each return series and JB Stat is the Jarque-Bera test statistic for the test of normality. ADF and PP are the respective test statistics for the Augmented Dickey-Fuller and Phillips-Perron unit root tests of stationarity.

Table 6.3: Serial Correlation Structure Of South African Stock Returns

	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	$Q(5)$	$Q(10)$	$Q^2(15)$	**ARCH(1)**	**ARCH(5)**	**ARCH(10)**
JSE All Share Index	-0.028	0.063	0.099	0.050	-0.080	4.623	7.237	62.586***	7.073***	3.009**	5.500***
Chemicals	-0.049	0.027	-0.037	0.164*	-0.055	7.027	9.413	31.817***	4.398**	1.605	1.486
Forestry & Paper	-0.007	-0.137	0.149*	0.190*	-0.071	16.235***	18.851**	18.851	0.000	2.729**	1.406
Ind. Metals & Mining	0.035	0.242*	-0.004	0.218*	-0.012	21.127***	23.098**	19.926	0.815	3.153***	1.556
Mining	-0.038	0.139*	0.046	0.037	-0.142*	8.808	12.073	23.918*	1.885	2.263*	2.066**
Const. & Materials	0.177*	0.099	-0.061	0.064	0.151	14.058**	20.855**	28.242**	11.057***	2.788**	2.226**
General Industrials	-0.031	-0.048	0.016	-0.146*	-0.032	5.1252	11.575	17.621	0.264	0.626	1.734*
Elec. & Elec. Equip.	0.147*	-0.090	0.091	0.032	0.023	7.749	18.490**	36.162***	0.554	5.654***	3.182***
Indust. Engineering	0.152*	0.007	0.005	0.140*	0.165*	13.860**	14.592	32.322***	0.089	3.903***	2.250**
Indust. Transp.	-0.000	0.106	-0.093	0.115	0.059	7.2055	12.418	24.142*	0.304	2.086*	1.488
Support Services	0.057	0.107	-0.087	-0.004	0.004	4.3767	7.605	20.821	0.720	0.746	0.557
Automobiles & Parts	0.117	-0.027	-0.052	0.083	0.084	6.095	18.430**	6.472	0.012	0.171	0.340
Beverages	-0.050	-0.11	-0.048	-0.008	-0.053	1.539	7.309	4.6764	0.199	0.145	0.443
Food Producers	0.057	-0.055	-0.018	-0.034	-0.053	2.077	2.722	12.447	0.001	2.061*	1.412
Health C. Equip & Serv.	0.050	0.021	0.026	0.165*	0.186*	13.039**	15.203	21.844	0.243	2.601**	1.582
Pharm & Biotech.	-0.014	-0.045	-0.011	0.103	-0.071	3.559	6.636	37.322***	0.174	0.288	2.558***
Food & Drug Retailers	-0.032	-0.025	-0.046	-0.102	0.048	3.228	5.106	13.679	0.064	0.400	0.370
General Retailers	0.105	-0.023	-0.003	-0.099	0.040	4.520	5.832	8.831	0.009	0.896	0.975
Media	0.025	-0.059	-0.141*	-0.054	0.215*	14.533**	24.733***	29.528**	1.812	1.501	4.217***
Travel & Leisure	0.103	0.033	-0.01	0.125	0.023	5.513	8.339	18.228	1.854	2.183*	1.446
Fix. Line Telecoms.	0.114	0.076	0.073	-0.022	0.162*	10.014*	15.840	20.983	0.200	1.769	1.714*
Banks	-0.070	-0.012	-0.093	0.038	-0.040	3.297	6.794	43.946***	2.794*	6.024***	3.236***
Non-life Insurance	0.019	-0.161*	-0.136	-0.009	0.051	9.367*	18.812**	19.205	0.364	1.973*	1.176
Life Insurance	0.013	0.013	0.019	0.074	0.114	3.813	9.041	51.191***	0.002	6.063***	4.532***
General Financial	0.132*	-0.053	0.031	0.066	0.132*	8.493	13.637	32.439***	3.637*	3.111***	1.695*
Equity Invest. Inst.	-0.091	0.135*	0.027	0.057	0.036	6.226	15.726	9.299	0.236	0.260	1.065
Software & Comp. Serv.	0.189*	0.045	0.036	0.081	0.094	10.690*	21.747**	77.734***	2.923*	6.530***	4.633***

Notes:

The asterisks, ***, ** and *, indicate statistical significance at the respective 1%, 5% and 10% levels of significance. Outliers are not excluded. ρ_1 to ρ_5 are the first five individual serial correlation coefficients for each return series. $Q(5)$ and $Q(10)$ are Ljung-Box Q-statistics indicating whether the first five and 10 serial correlation coefficients are jointly equal to zero. $Q^2(15)$ is the Ljung-Box test statistic for non-linear dependence in returns up to the 15th order. ARCH(1), ARCH(5) and ARCH (10) are LM test statistics for ARCH effects in the residuals of the AR(1) model in equation (6.13) at the first, fifth and 10th orders respectively.

The observed departures from non-normality and the presence of non-linear dependence in returns are indicative of time-varying variance. The presence of ARCH effects in the residuals of an AR(1) model indicates that the least squares methodology may not be the most appropriate estimation methodology.

The broad factor set identified on the basis of the dividend discount model (equation (6.2)) is presented in Table 6.4. The statistical properties of the respective factors are reported for the differenced and pre-whitened factor series. The factors listed reflect a mixture of domestic and global macroeconomic factors in line with findings that show that the South African economy is partially integrated with the global economy (Szczygielski & Chipeta, 2015: 3, 16).

Following pre-whitening, almost all factor series represent zero-mean series of innovations. The exceptions are changes in the terms of trade (TOT_t) and the inflation rate (CPI_t). The Breusch-Godfrey LM test indicates that the factor series derived from equation (6.3) for both factors are serially correlated up to 12 lag orders although the respective *Q*-statistics for these series are statistically insignificant. The remaining factors meet the requirements of the APT; all series are characterised by a mean value of zero and are pure innovations. This is confirmed by statistically insignificant mean values (as established by applying the *t*-test to test the null hypothesis that $E(f_{kt}) = 0$), insignificant Breusch-Godfrey LM test statistics and *Q*-statistics for 12th order serial correlation. The autoregressive time series methodology employed to generate innovations yields highly satisfactory results and appears to adequately represent the expectations generating process.

Table 6.4: Factor Set

Factor	Form	Notation	Mean	Std Dev.	LM Test	$Q(12)$	Lags	ADF	PP
Panel A: Real Activity									
Manufacturing Sales	DL	MFS_t	-0.000	0.022	1.167	6.702	1,6,12	-14.295***	-14.304***
Wholesale Trade sales	DL	WHL_t	-7.52E-05	0.024	0.887	10.602	1	-14.056***	-14.107***
Retail Trade Sales	DL	RET_t	-0.000	0.012	0.758	4.507	1,6,9,11,12	-13.821***	-13.823***
New Vehicle Sales	DL	VEH_t	-0.000	0.043	1.010	10.153	1, 7-8	-14.071***	-14.113
Total Mining Prod.	DL	MIP_t	2.31-E05	0.037	0.923	8.967	1-4	-14.090***	-14.164***
Building Plans Passed	DL	BP_t	-0.002	0.108	1.156	8.885	1-2, 4	-14.418***	-14.475***
Buildings Completed	DL	BC_t	0.000	0.156	0.492	5.258	1-5	-13.518***	-13.517***
Employment	DL	EMP_t	-0.000	-0.058	1.630	15.316	1-2	-13.723***	-13.733***
Panel B: Prices									
Consumer Price Inflation	PC	CPI_t	-2.45E-05	0.004	2.059**	5.628	1,3, 12	-12.710***	-12.751***
Inflation Expectations	D	BAR_t	-4.88E-06	0.003	0.684	7.857	1-3, 12	-13.576***	-13.597***
Prod. Price Index	PC	PPI_t	-0.000	0.082	0.384	3.316	1-2, 11	-13.764***	-13.776***
Input Prices	PC	INP_t	0.000	0.062	1.489	18.038	-	-13.543***	-13.916***
Panel C: Cyclical Indicators									
Inventories	DL	INV_t	0.000	0.077	0.883	7.086	1-2, 12	-14.099***	-14.126***
Leading Indicator	DL	$LEAD_t$	-3.07E-05	0.008	1.036	6.536	1-3, 6, 11-12	-13.615***	-13.618***
Coincident Indicator	DL	$COINC_t$	-2.68E-06	0.005	1.196	8.735	1-2, 6-7, 11	-13.931***	-13.931***
Lagging Indicator	DL	LAG_t	-3.63E-05	0.009	11.943	11.536	1-2	-13.843***	-13.843***
House Prices	DL	HSE_t	9.66E-06	0.031	1.22	16.752	1-2, 9	-13.708***	-13.813***
Business Activity	DL	BUS_t	0.000	0.087	0.583	5.071	1 – 2, 7, 9,	-14.501***	-14.568***
Panel D: Exchange Rates									
Rand-Dollar Ex. Rate	DL	USD_t	-0.001	0.036	0.609	8.168	1,8	-13.522***	-13.518***
Rand-Euro Ex. Rate	DL	EUR_t	-0.000	-0.035	1.001	11.784	1, 5, 8	-13.787***	-13.799***
Rand-Pound Ex. Rate	DL	GBP_t	-0.000	0.036	0.824	9.240	1, 3, 8	-13.397***	-13.390***
Nominal Effective Ex. Rate	DL	NEX_t	0.000	0.033	0.906	10.983	1, 5, 8	-13.755***	-13.764***
Real Effective Ex. Rate	DL	REX_t	0.000	0.033	0.616	5.338	1-2, 5, 8	-13.961***	-13.977***

Table 6.4: Factor Set (Continued…)

Panel E: Monetary Factors									
M0 Monetary Aggregate	DL	$M0_t$	-0.000	0.016	0.450	5.4932	1	-14.552***	-14.525***
M1A Monetary Aggregate	DL	$M1A_t$	0.001	0.024	0.436	4.709	1-2, 7, 11-12	-14.224***	-14.219***
M1 Monetary Aggregate	DL	$M1_t$	0.001	0.022	1.134	6.940	1-2, 12	-13.250***	-13.264***
M2 Monetary Aggregate	DL	$M2_t$	0.000	0.013	0.943	6.226	8, 12	-14.576***	-14.577***
M3 Monetary Aggregate	DL	$M3_t$	0.000	0.010	0.966	11.617	2-3, 12	-14.256***	-14.263
Excess M3 Supply Growth	DL	$M3E_t$	-0.001	-0.049	1.697*	13.283	1, 12	-15.320***	-15.287***
Total Credit Extension	DL	TCR_t	-1.33E-06	-0.013	0.779	8.081	1, 3, 12	-14.312***	-14.405***
Private Credit Extension	DL	PVC_t	-2.16E-05	0.011	1.263	10.319	2-3, 7	-13.604***	-13.617***
Gold Reserves	DL	GFR_t	-0.003	0.068	0.670	7.345	1, 11-12	-14.417***	-14.432***
Foreign Reserves (US$)	DL	RES_t	0.001	0.022	1.411	6.038	1,3, 6 8 -10	-13.701***	-13.726***
Foreign Reserves (Rand)	DL	$RESZ_t$	0.000	0.037	0.588	5.549	1, 8	-14.018***	-14.021***
Panel F: Commodities									
Commodities (US$)	DL	COM_t	0.000	0.046	0.918	9.511	1, 6, 10	-14.068***	-14.082***
Commodities (Rand)	DL	$COMZ_t$	-0.001	0.049	0.662	5.262	1, 6	-13.550***	-13.552***
Non-fuel Commodities	DL	NFC_t	0.000	0.026	0.689	3.782	1, 4, 8, 11	-13.973***	-13.981***
Non-fuel Commodities	DL	$NFCZ_t$	0.000	0.038	0.555	8.979	1	-13.377***	-13.369***
Oil Prices (US$)	DL	OIL_t	6.51E-05	0.087	1.180	13.962	1, 6	-13.212***	-13.207***
Oil Prices (Rand)	DL	$OILZ_t$	-0.001	0.085	1.222	14.085	1	-13.181***	-13.173***
Gold Prices (US$)	DL	GLD_t	0.000	0.038	0.671	6.083	1, 4, 11	-13.314***	-13.302***
Gold Prices (Rand)	DL	$GLDZ_t$	-0.000	0.046	0.885	9.829	1, 4	-13.331***	-13.331***
Metal Prices (US$)	DL	MET_t	0.000	0.047	0.678	6.324	1	-13.752***	-13.759***
Metal Prices (Rand)	DL	$METZ_t$	0.000	0.048	0.452	5.481	1	-13.337***	-13.289***
Panel G: Interest Rates									
Real Interest Rates	DL	RIB_t	0.001	0.022	0.767	6.478	1, 12	-13.854***	-13.854***
3-Month T Bill Rates	D	$3TB_t$	2.59E-05	0.003	1.111	8.635	1, 4, 8	-14.223***	-14.228***
Long-Term Gov. Bond Yields	D	LTY_t	3.97E-05	0.003	0.990	8.453	1-2, 7	-13.935***	-13.941***
Short-Term Gov. Bond Yields	D	STY_t	6.28E-05	0.004	0.672	3.641	1, 7	-13.859***	-13.859***
Term Structure	D	TER_t	1.71E-05	0.003	0.834	9.832	1, 3	-13.711***	-13.711***

Table 6.4: Factor Set (Continued…)

Panel H: Trade									
Trading Partner Lead. Index	DL	TLI_t	9.30E-05	0.005	0.568	7.938	1, 3, 6	-13.516***	-13.518***
Trading Partner Coinc. Index	DL	TCI_t	-3.14E-05	0.002	1.325	8.898	1, 11-12	-15.029***	-15.031***
Terms of Trade	DL	TOT_t	1.23E-05	0.041	3.356***	13.421	1-12	-14.276***	-14.276***
Panel I: Market Indices									
JSE All Share Index	DL	R_{Mt}	0.005	0.048	0.397	7.956	-	-13.991***	-14.056***
MSCI World Index (US$)	DL	R_{IMt}	0.001	0.044	0.923	7.986	1,3	-13.363***	-13.355***
MSCI World Index (Local)	DL	R_{IMLt}	0.000	0.041	1.005	13.584	1	-14.667***	-14.643***
MSCI World Index (Rand)	DL	R_{IMZt}	0.000	0.047	0.913	9.594	3, 10	-14.565***	-14.574***

Notes:
The asterisks, ***, ** and *, indicate statistical significance at the respective 1%, 5% and 10%% levels of significance. Form indicates the method of differencing used to derive changes in a given factor, where FD=First Difference, FDL=First Logarithmic Differences and PC=Percentage Changes. Notation refers to the formulaic notation used to abbreviate each factor. Mean and Std Dev. are the respective mean and standard deviation values for each factor series. LM Test is the Breusch-Godfrey LM test statistic for 12th order serial correlation in a factor series. $Q(12)$ are Ljung-Box *Q*-statistics indicating whether the first 12 serial coefficients for a given factor series are jointly equal to zero. Lags indicates the lag orders that are retained in the autoregressive model in equation (6.3) used to derive innovations in the factor series. ADF and PP are the respective test statistics for the Augmented Dickey-Fuller and Phillips-Perron unit root tests of stationarity.

6.3. FACTOR STRUCTURE AND THE RESIDUAL MARKET FACTOR

6.3.1. Factor Analysis And The Factor Structure

The exploration of the factor structure of the returns series comprising the sample begins with a summary of the correlation matrix of the return series. This is done by reporting a histogram, relative frequencies, minimum and maximum levels of across series correlation and the mean level of correlation across the return series in the sample. The aim of this analysis is to gain an insight into the nature and strength of interdependence between return series. Underlying this analysis is the assumption that interdependence is attributable to co-movement in returns in response to systematic factors (Elton & Gruber, 1988: 31; Elton *et al.,* 2014: 157). The correlation matrix of the return series is also reported and later on, the histogram, extreme values, the mean correlation and the correlation matrix are used in comparisons with residuals derived from models applied to investigate the impact of factor omission and the ability of the residual market factor to resolve underspecification. As an informal and limited test of the significance of pairwise return correlations, the aggregated return correlation coefficients are tested against a null hypothesis of a mean and median of zero using the *t*-test and the Wilcoxon test (Eichholtz, 1996: 61; Section 6.4.9.). As there may be ambiguity regarding the distribution of the coefficients, the non-parametric Wilcoxon matched-pairs signed-rank test is applied, as correlation coefficients may not be normally distributed. The null hypothesis that is tested in this test, is that the median does not differ significantly from zero (Chen & Jordan, 1993: 80; Artiach, Lee, Nelson & Walker, 2010: 39, 42).

In the next step, factor analysis is employed to investigate the pervasive influences in stock returns, which are represented by statistical factors derived using factor analysis (Connor, 1995: 42). Factor analysis has a long history of application in APT literature and early studies apply factor analysis to derive the number of factors in the linear factor model, factor scores and the associated factor loadings (betas) that fulfil the role of inputs in the APT relation (Section 2.2.). Factor analysis permits a derivation of factors that explain common variance, namely the communality, from a larger dataset and aims to find the smallest number of common factors that account for correlations (McDonald, 1985; Yong & Pearce, 2013: 80, 82). In the context of the APT, this is the proportion of variance that is explained by pervasive influences that are common to all return series. Other types of variance that require further mention are uniqueness and error variance. Uniqueness is the variance that is specific to a given return series and is therefore not systematic. Error variance is the variance that is attributable to random components or systematic error in the model (Walker & Madden, 2008: 326-237). An example of the application of factor analysis to explore the structure of the residual correlation matrix of the macroeconomic linear factor model can be found in Van Rensburg (1997: 63) (also see Section 3.4.).

To gain a preliminary insight into the factor structure of South African stock market returns, a scree test is first conducted on the correlation matrix of the return series in the sample. This test yields a scree plot and the resultant flexion point, which demarcates a steep gradient (systematic factors) from a flat gradient (trivial factors), is indicative of the number of true systematic factors (Kryzanowski & To, 1983: 37). In the next step, the factor structure is investigated by estimating the associated eigenvalues and the percentage of variance accounted by the first 10 factors (Hughes, 1984: 207). Although the scree test finds application in APT literature, Yong and Pearce (2013: 85) warn that this test is only reliable for a sample size of at least 200 (the current sample consists of 26 sectors). Therefore, the minimum average partial (MAP) test is also applied. The MAP test, introduced by Velicer (1976), seeks to extract the best factor solution as opposed to finding a cut-off point for the number of factors (Ledesma & Valero-Mora, 2007: 3).[92] This approach is appealing; the MAP test seeks to derive the number of factors that result in a minimum average squared partial correlation and a solution is reached when the residual matrix most closely resembles an identity matrix (Zwick & Velicer, 1986: 434). In other words, off-diagonal correlations are zero or close to zero, in line with the assumption of uncorrelated residuals, $E(\varepsilon_{it}, \varepsilon_{jt}) = 0$ (equation (2.2)). Importantly, this approach extracts factors that explain common variance as opposed to residual or error variance (Courtney, 2013: 3).

In the final step, the factor scores are derived using the Bartlett (1937) method. This method produces factor scores that are unbiased estimates of the true factor scores and are most likely to represent the true factor scores as scores are estimated using ML estimation (Bartlett, 1937; DiStefano, Zhu & Mîndrilă, 2009: 4-5). To facilitate a clearer interpretation of the factors at a later stage (Section 6.3.2.), the extracted factors are subjected to an orthogonal varimax rotation. The purpose of rotation is to have return series load on as few factors as possible but to maximise the number of high loadings. The varimax rotation procedure minimises the number of return series that have high loadings on each factor

[92] It has been shown that this method is accurate under many conditions (EViews 7 User's Guide II, 2009: 709; Ledesma & Valero-Mora, 2007: 3).

(thus reducing the number of factors with high loadings) and reduces small loadings further (Abdi, 2003: 3; Yong & Pearce, 2013: 84). Factor scores are retained for use in the selection of macroeconomic factors that are assumed to be proxies for the pervasive influences represented by the extracted factors.

6.3.2. Factor Selection

The next step is to select macroeconomic factors, represented by the macroeconomic factor innovation series, from the broader factor set outlined in Table 6.4., which will feature in the linear factor models used to investigate the consequences of underspecification and the role of the residual market factor. Therefore, it is necessary to identify the factors that are the best proxies for the pervasive influences in returns. Although the literature recognises that there is no unique set of factors that explain returns, it is desirable to use a set of factors that will capture the influence of other factors that are also relevant but not included in the model (McElroy & Burmeister, 1988: 41). By selecting this set of factors, it is hoped that as much as possible of the systematic variation in returns will be reflected by this factor set.

The identification of factors begins with establishing which factors are systematically correlated with stock returns. As macroeconomic data is often subject to revisions and/or lags in announcements, for example, January inflation figures may only be announced in the following month and therefore returns respond to January's inflation in February, factors must enter the correlation matrix in a manner that reflects lags in announcements (Clare & Thomas, 1994: 313).[93] Consequently, each factor enters the factor-return correlation matrix contemporaneously and with three lags (Bilson et al., 2001: 406; Panetta, 2002: 424). A summary of the correlation of each factor and the lag order that is significantly correlated with the greatest number of industrial sectors and the JSE All Share Index is reported (Section 7.2.). The primary form of correlation that is estimated is Pearson's (ordinary) correlation. As ordinary correlation may be unreliable in the presence of non-normality, heteroscedasticity and outliers, the presence of significant correlation (or lack thereof) between a given factor and the JSE All Share Index is confirmed using non-parametric Spearman's (rank) correlation coefficients (Onwuegbuzie & Daniel, 2002; Bonnet & Wright,

[93] For example, Statistics South Africa, South Africa's governmental statistical agency, releases retail trade sales figures with a 45 day delay. Therefore, returns at time *t* may respond to innovations in this factor between time *t-2* and *t-3.* See Statistics South Africa's (2017) statistical release P6242.1.

2000: 24; Hauke & Kossowski, 2011: 89; Bishara & Hittner, 2012: 406).[94] Only factors that are significantly correlated with at least half of the return series in the sample and returns on the JSE All Share Index are considered as candidate factors for inclusion in the linear factor model. In instances where factors are correlated with more than half of the return series but not with returns on the JSE All Share Index, the Bai-Perron (1998) test is applied to determine the existence and number of breakpoints and to identify breakpoints, if these exist. Single-factor breakpoint least squares models are estimated to confirm the absence of a relationship by regressing returns on the JSE All Share Index onto the factor in question (Hansen, 2012: 76).[95] This is to ensure that an apparent lack of correlation is not driven by structural changes in the relationship between aggregate returns and a specific factor. By establishing that factors are correlated with returns on the JSE All Share Index and at least half of the return series, it is hoped that the factors taken forward for further analysis are truly pervasive in nature as opposed to being pseudofactors (Kryzanowksi & To, 1983: 39; Connor, 1995: 44).

The next step brings together the results of the identification of the number of factors, the extraction of factor scores and the results of the factor-return correlation analysis. Relationships are established between the factor scores for factors extracted using factor analysis and the set of macroeconomic factors identified in the factor-return correlation analysis. In this step, proxy factor regressions are estimated by regressing factor scores onto the abovementioned joint (identified) set of factors. To establish whether the macroeconomic factors retained for further analysis are proxies for the statistical factors that represent pervasive influences in returns, the approach of Chen and Jordan (1993: 73), Choi and Rajan (1997: 42-43), Panetta (2002: 430-433) and Spyridis *et al.* (2012: 52) is adapted for the purposes of the present study (also see Hahn & Lee, 2006; Aretz *et al.,* 2010).[96] The

---

[94] The use of an additional measure of correlation is motivated by a preliminary observation that in some instances, a factor is significantly correlated with a number of industrial sectors but uncorrelated (often marginally insignificant) with the JSE All Share Index. This, at times, contradicts prior findings. Given the limitations of correlation analysis, the results reported in Section 7.3. should be seen as indicative of possible relationships but not as definitive evidence. For examples of the use of Spearman's correlation with economic data, see Liow, Ibrahim & Huang (2006: 305) and Naifar and Al Dohaiman (2013: 428).

[95] Model settings used are: L+1 vs. L. sequentially determined breaks, Trimming: 0.15, Max. Breaks: 5, Sig. level: 0.05 with HAC standard errors and covariance heterogeneous distributions across breakpoints assumed. The results of single-factor breakpoint least squares regressions are not reported but are available upon request.

[96] The studies of Hahn and Lee (2006) and Aretz *et al.* (2010) investigate whether the Fama-French factors (e.g. book-to-market (HML), size (SML), momentum factors (MOM)) serve as proxies for macroeconomic factors. The principle is similar.

retained macroeconomic factors, are regressed onto the factor scores of each statistical factor, F_{nt}:

$$F_{nt} = \alpha + \sum_{k=1}^{K} b_{nk} f_{kt} + \varepsilon_{nt} \tag{6.14}$$

where b_{nk} is the sensitivity of statistical factor *n* to macroeconomic factor *k.* It must be emphasised that factors derived from factor analysis represent a function of factors and are not factors by themselves (Rummel, 1967: 459). It follows that if macroeconomic factors are significantly related to at least some of the statistical factors in this multifactor analysis, then these factors are proxies for pervasive influences in stock returns (Panetta, 2002: 430). Such factors are retained for incorporation into the specifications of the linear factor model that follow.

This study seeks to determine not only the adequacy of the macroeconomic APT linear factor model but also the ability of a residual market factor to resolve underspecification. Therefore, the next step involves incorporating the conventional residual market factor into equation (6.14). To derive the residual market factor, returns on the JSE All Share Index are regressed onto the identified macroeconomic factor set, as follows:

$$R_{Mt} = \alpha + \sum_{k=1}^{K} b_{Mk} f_{kt} + M\varepsilon_t \tag{6.15}$$

where R_{Mt} are the returns on the JSE All Share Index at time *t*, b_{Mk} is the sensitivity of R_{Mt} to innovations in macroeconomic factor *k* and $M\varepsilon_{Mt}$ are now the residuals in equation (6.15). The residual market factor, $M\varepsilon_t$, is now incorporated into equation (6.14), as follows:

$$F_{nt} = \alpha + \sum_{k=1}^{K} b_{nk} f_{kt} + b_{nM\varepsilon} M\varepsilon_t + \varepsilon_{nt} \tag{6.16}$$

where all parameters are as in equation (6.14), but $M\varepsilon_t$ is the residual market factor derived in equation (6.15) and the $b_{nM\varepsilon}$ is the coefficient on $M\varepsilon_t$. The inclusion of $M\varepsilon_t$ in equation (6.16) not only tests the ability of the residual market factor to proxy for pervasive influences but also tests whether the macroeconomic factors by themselves are adequate proxies. If all influences that impact returns are reflected in the macroeconomic factor set, then $b_{nM\varepsilon}$ in equation (6.16) will be insignificant. As the residual market factor is orthogonal to the set of macroeconomic factors by construction, the inclusion of this factor will have no impact on any of the factor coefficients in equation (6.16) as this specification (and the other factor-proxy regressions) is estimated using least squares estimation (Wurm & Fisicaro, 2014: 31; Czaja *et al.,* 2010: 130). It is expected that the residual market factor will be statistically significant as macroeconomic factors are unlikely to account for all pervasive influences in returns. As a final step, equation (6.16) is extended to include the second residual market factor, derived in equation (6.17). This residual market factor is derived from returns on the MSCI World Market Index, expressed generically as R_{It} at this stage but denominated either in US Dollars (R_{IMt})[97] or in local currency terms (R_{IMLt}) or in Rand terms (R_{IMZt}), as follows:[98]

$$R_{It} = \alpha + \sum_{k=1}^{K} b_{Ik} f_{kt} + b_{It} R_{Mt} + I\varepsilon_t \tag{6.17}$$

where R_{It} is the return on the MSCI World Market Index (denominated in US Dollars, or the local currency or in Rands), b_{Ik} is the sensitivity of returns on the MSCI World Market Index to factor *k*, f_{kt}, and b_{It} is the sensitivity to returns on the JSE All Share Index, R_{Mt}.[99] $I\varepsilon_t$ is the residual series of this specification – the second residual market factor. This way, the international residual market factor represents all influences that are not reflected in returns on the JSE All Share Index (from which $M\varepsilon_t$ is derived) and the macroeconomic factor set.

The second residual market factor, $I\varepsilon_t$, is now incorporated into equation (6.18), so that:

---

97 The derivation of excess returns on the Dollar denominated MSCI World Index closely follows the methodology in Ferson and Harvey (1994: 798). Rates on the 90-day US treasury bill are used as a proxy for the risk-free rate, as opposed to a one-month treasury bill rate, which is unavailable. Excess returns on the local currency MSCI World Index are derived by estimating month-to-month returns and then subtracting the rate on the 90 day US treasury bill. This approach is also taken with the MSCI World Market Index denominated in Rands. A more appropriate approach would be to use rates calculated using 90-day US treasury bills denominated in Rand. However, historical price data for 90-day US Dollar treasury bill auction was unavailable to the author.

98 The local currency MSCI World Market Index, as constructed by MSCI, represents only the changes in the prices of stocks constituting the MSCI World Market Index and excludes the impact of currency fluctuations (MSCI Index Calculation Methodology, July 2017).

99 It is assumed that causality runs from the MSCI World Market Index to the JSE All Share Index. Therefore, b_{It} is a measure of the strength of the relationship between returns on the two indices and does not have a further interpretation. The same may be said about macroeconomic factors that reflect domestic economic conditions and do not impact global market returns.

$$F_{nt} = \alpha + \sum_{k=1}^{K} b_{nk} f_{kt} + b_{nM\varepsilon} M\varepsilon_t + b_{nI\varepsilon} I\varepsilon_t + \varepsilon_{nt} \tag{6.18}$$

where $I\varepsilon_t$ is a generic representation of the second residual market factor, b_{nk} is the sensitivity of factor *n* to innovations in macroeconomic factor *k,* $b_{nM\varepsilon}$ is the sensitivity to the conventional residual market factor, $M\varepsilon_t$, and $b_{nI\varepsilon}$ is the sensitivity to the second residual market factor, $I\varepsilon_t$, derived from a version of MSCI World Market Index (see above). It follows that if the macroeconomic factor set and the domestic residual market factor are adequate proxies for the underlying common factors, the international residual market factor will be redundant (Chang, 1991: 380; Kryzanowski *et al.,* 1994: 155-156). This can be seen as a preliminary test of the adequacy of the residual market factor as a proxy for omitted factors.

The estimation of equation (6.14), the subsequent inclusion of $M\varepsilon_t$ in equation (6.16) and the inclusion of $I\varepsilon_t$ in equation (6.18) permits a confirmation and comparison of the ability of the macroeconomic factor set and the residual market factors to proxy for pervasive influences in returns. This approach presents an early foray into the investigation of underspecification as it indicates whether the macroeconomic factors and the residual market factors approximate the pervasive influences in stock returns. The resultant $\bar{R}^2$ s are treated as measures of how well the included factors reflect the underlying pervasive influences represented by the statistical factors (Panetta, 2002: 430; Aretz *et al.,* 2010: 1388). Consideration is also given to the Akaike Information Criterion (AIC), informing which combination of factors best explains the factor scores and therefore yields the best approximation of underlying pervasive factors in the South African stock market (Mills & Markellos, 2008: 34, 231; Spiegelhalter, Best, Carlin & Van der Linde, 2014: 1-2). The *F*-test is applied to test the joint significance of the coefficients and to confirm the overall significance of the model. This is essentially a test of the null hypothesis, namely that the $\bar{R}^2$ is zero, and a rejection of this hypothesis confirms the significance of a given factor set in approximating the statistical factors (Sadorsky, 2001: 25; Gujarati & Porter, 2009: 111).

6.4. UNDERSPECIFICATION IN THE LINEAR FACTOR MODEL

6.4.1. Approach And Model Specification

The approach undertaken in this study is comparative; four specifications are estimated and compared to establish the consequences of underspecification on various aspects of the linear factor model. The comparative aspect lies in the comparison of the results across the different specifications and the quantification of changes in the numerous parameters and aspects of these specifications as factors are included and omitted.

The first specification is a benchmark model, which is hypothesised to be free of underspecification attributable to the omission of systematic factors. This specification is estimated in three steps. In the first step, the specification follows the functional form of equation (6.18) and the dependent factor is now the series of returns on industrial sector *i* at time *t*, R_{it}. In this step, each industrial sector return series is regressed onto the (by now identified) macroeconomic factor set and the two residual market factors, $M\varepsilon_t$ and $I\varepsilon_t$:

$$R_{it} = \alpha + \sum_{k=1}^{K} b_{ik} f_{kt} + b_{iM\varepsilon} M\varepsilon_t + b_{iI\varepsilon} I\varepsilon_t + \varepsilon_{it} \tag{6.19}$$

where $\sum_{k=1}^{K} b_{ik} f_{kt}$ is the macroeconomic factor set consisting of innovations in the identified macroeconomic factors and the associated sensitivities to innovations in the respective factors, b_{ik} s. $M\varepsilon_t$ is the conventional residual market factor, $b_{iM\varepsilon}$ is the associated coefficient, $I\varepsilon_t$ is the international residual market factor and $b_{iI\varepsilon}$ is the associated coefficient. The residual terms for each sector are denoted by ε_{it}. Equation (6.19) represents the unrestricted model, which is postulated to be initially fully specified as it includes the two residual market factors which are expounded in the literature to proxy for any omitted domestic and international factors (Section 3.2.; Clare & Priestley, 1998: 110-111). McElroy and Burmeister (1988: 41) argue that although there may be other factors and that there is no unique set of factors, all (macroeconomic) influences, even if more important factors exist than the factors included in the unrestricted model, will be captured by the macroeconomic factor set identified and a residual market factor (or factors as in this specific study).[100] According to this line of reasoning, the unrestricted model comprising the macroeconomic

[100] The reader is reminded that consideration is given to the ability of the macroeconomic factors to proxy for the statistical factors representative of systematic influences in stock returns in Chapter 7. Specifically, the $\bar{R}^2$ is considered and factors with the best ability to approximate the statistical factors are chosen as proxies for the factor scores. As such, out of the extensive set of macroeconomic factors screened and taken forward for analysis, these factors are not only systematic in nature but also the best available proxies for underlying systematic factors.

factors and $M\varepsilon_t$ and $I\varepsilon_t$ should account for all systematic influences and should not be underspecified, even if the optimal set of factors is not used.

The estimation of the unrestricted specification in equation (6.19) is the first step in constructing the benchmark model. As inference is not of interest at this stage of the analysis and all that is of interest is the derivation of the residuals, ε_{it}, equation (6.19) is estimated using the least squares methodology. The residuals are central to the second step. If the residual series are correlated with omitted factors and consequently there is correlation across the residual series (Section 3.2., equation (3.3)), the residuals can be decomposed as follows (restated) (McElroy & Burmeister, 1988: 33; Burmeister & McElroy, 1991: 43):

$$\varepsilon_{it} = +\sum_{j=1}^{J} b_{ij} f_{jt} + \varepsilon_{it}^{*} \tag{3.3}$$

where ε_{it} is the residual term in equation (6.19) for each series, $\sum_{j=1}^{J} b_{ij} f_{jt}$ is the set of omitted observed and unobserved common factors and the associated coefficients and ε_{it}^{*} is the idiosyncratic residual term that now no longer reflects any omitted factors. In the presence of omitted and unobserved factors and for the purposes of estimation, the f_{jt} s are replaced with well-diversified portfolios, which are likely to be (but not necessarily) aggregate market indices (Burmeister & McElroy, 1991: 39; 44). As equation (6.19) already includes the residual market factors that are of interest in this study, f_{jt} should not exist in equation (3.3) if the residual market factors account for all omitted factors. The omission of systematic factors in the residuals of equation (6.19) can therefore be established by factor analysing the resultant correlation matrix comprising the residuals for each industrial sector. It then follows, in the second step, that the resultant residual correlation matrix is factor analysed in the same manner as returns in Section 6.3.1. to derive the common omitted and unidentified underlying factors in the residual series of the industrial sectors. If systematic factors are still present in the residual correlation matrix, then in the third step, the unrestricted model in equation (6.19) is augmented with the derived factors to account for the omitted and unobserved common factors (Van Rensburg, 1997: 61, 63; 2000: 36-37):

$$R_{it} = \alpha + \sum_{k=1}^{K} b_{ik} f_{kt} + b_{iM\varepsilon} M\varepsilon_t + b_{iI\varepsilon} I\varepsilon_t + \sum_{j=1}^{J} b_{ij} f_{jt} + \varepsilon_{it}^{*} \tag{6.20}$$

where all parameters of the model are as before as in equation (6.19), with the exception of $\sum_{j=1}^{J} b_{ij} f_{jt}$ which is the set of derived factors and associated sensitivities and ε_{it}^{*} represents purely idiosyncratic factors devoid of systematic components. Van Rensburg (2000: 37) states that this approach “can be viewed as an econometric correction for omitted variable bias,” a factor analytic augmentation. Equation (6.20) is the benchmark model; a model that is assumed to be free of underspecification and which serves as a specification against which the restricted and unrestricted specifications are compared. This specification is discussed in detail in Chapter 8.

The next step is to specify a restricted model, which is potentially underspecified and incorporates only the macroeconomic factor set:

$$R_{it} = \alpha + \sum_{k=1}^{K} b_{ik} f_{kt} + \varepsilon_{it} \tag{6.21}$$

where $\sum_{k=1}^{K} b_{ik} f_{kt}$ consists of the identified and retained macroeconomic factors that proxy for the pervasive influences in stock returns and the associated coefficients. Various parameters and aspects of equation (6.21) (discussed in Section 6.4.4. to Section 6.4.9.) are compared to those of the benchmark model and the results of this comparison are reported in Chapter 9. The aim of this is to analyse the impact of factor omission, namely the two residual market factors and the factor analytic augmentation that are now omitted, on various aspects of the specification.

Two further specifications are estimated and each is compared to the benchmark model and the restricted specification set out above. The first is the unrestricted market model, which incorporates the domestic residual market factor, $M\varepsilon_t$:

$$R_{it} = \alpha + \sum_{k=1}^{K} b_{ik} f_{kt} + b_{iM\varepsilon} M\varepsilon_t + \varepsilon_{it} \tag{6.22}$$

where $\sum_{k=1}^{K} b_{ik} f_{kt}$ is the set of macroeconomic factors and associated coefficients and $M\varepsilon_t$ is the residual market factor derived from returns on the JSE All Share Index. The comparison of the results for equation (6.22) with those of the benchmark model in equation (6.20) and the restricted model in equation (6.21) seeks to determine whether the inclusion of the residual market factor improves various aspects of the unrestricted market model relative to the restricted model and whether a specification that includes the conventional residual market factor approximates the fully specified benchmark model.

The final specification is the unrestricted model denoted by equation (6.19). The results and various aspects of this model are compared to those of the benchmark model and the restricted model and also, partly, to those of the unrestricted market model. The purpose of this latter comparison is to establish whether the inclusion of a second residual market factor, $I\varepsilon_t$, yields a better approximation of the fully specified benchmark model relative to a specification that incorporates only the conventional residual market factor.

6.4.2. Econometric Methodology

The "workhorse" of applied econometrics, as Engle (2001: 157) refers to the least squares model, is the natural and first choice of econometric methodology for estimating equations (6.19), (6.20), (6.21) and (6.22). The least squares methodology is well understood, does not require the specification of a conditional error distribution and is easily implemented using readily available and accessible statistical software. A drawback is that in the presence of heteroscedasticity, coefficient estimates are inefficient and standard errors may be overestimated, resulting in misleading inferences (Engle, 2001: 157; Sadorsky, 2001: 203).

An approach to handling the volatility dynamics and serial correlation inherent in the residuals of a least squares regression is to use robust standard errors for parameter estimates such as the Newey and West (1987) heteroscedasticity and autocorrelation consistent (HAC) covariance matrix (HAC standard errors) or White's (1980) heteroscedasticity-corrected standard errors (White standard errors). However, the use of robust standard errors does not yield the most efficient coefficient estimates and the heteroscedasticity consistent covariance matrix estimator may be biased under certain conditions (Chesher & Jewitt, 1987; Andersen *et al.,* 2003: 48). A further important drawback of the least squares methodology is that when orthogonalised factors are included in the model, notably $M\varepsilon_t$ and $I\varepsilon_t$, the coefficients on the factors already in the model will not adjust to reflect the reduction in potential bias attributable to the inclusion of previously omitted factors (Dominguez, 1992: 95). Efficiency and bias are key aspects considered in this study as coefficient inefficiency and bias are consequences of underspecification (Section 5.3.1.). Therefore, it is desirable to have the most efficient coefficient estimates possible and for bias attributable to underspecification in the restricted specifications and unrestricted specifications, if present, to be reflected in model coefficients. This permits for a quantification of underspecification induced inefficiency in the standard errors and bias in model coefficients, allowing for comparisons of the restricted and unrestricted models to the benchmark specification, which is hypothesised to be free from underspecification.

The loss of efficiency and the quantification of the bias associated with underspecification can be achieved by applying the ARCH/GARCH framework. The ARCH/GARCH methodology incorporates the conditional variance into the log-likelihood function, which is used in the estimation of model parameters under ML estimation. The log-likelihood function, assuming conditional normality, is given by (Bera *et al.,* 1988: 206; Herwartz, 2004: 202):

$$L = \sum_{t=1}^{T}\left(-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\sigma_{it}^{2} - \frac{1}{2}\frac{\varepsilon_{it}^{2}}{\sigma_{it}^{2}} \right) \tag{6.23}$$

where $-\frac{1}{2}\log(2\pi)$ is a constant and ε_{it}^{2} are the squared residuals of a (multifactor) specification such as equations (6.19), (6.20), (6.21) and (6.22). ML estimation maximises L, the log-likelihood function, by estimating model parameters, the intercept (α) and the betas, b_{ik} s, that minimise ε_{it}^{2}. Assuming that $\sigma_{it}^{2} = h_{it}$, the conditional variance can be described by an ARCH(*p*) or GARCH(*p,q*) process (or any other type of ARCH/GARCH model in general) as follows:

$$h_{it} = \omega + \sum_{i=1}^{p}\alpha_i\varepsilon_{it-p}^{2} \tag{6.24}$$

$$h_{it} = \omega + \sum_{i=1}^{p}\alpha_i\varepsilon_{it-p}^{2} + \sum_{i=1}^{q}\beta_i h_{it-q} \tag{6.25}$$

where h_{it} is the conditional variance underlying return series *i*, ω is the unconditional variance, ε_{it-1}^{2} are the squared residuals conditional on model specification and h_{it-q} is the previous forecast of the conditional variance and its associated GARCH coefficient, β_i (Engle, 2004: 412). The number of ARCH terms is denoted by *p* and the number of GARCH terms is denoted by *q*. Of particular importance is α_i, the ARCH coefficient, which Bera *et al.* (1988: 204) treat as a measure of conditional heteroscedasticity. The authors provide early evidence that conditional heteroscedasticity, heteroscedasticity dependent upon the factors in a model, has an impact on coefficient estimates. The beta coefficient estimates of a market model estimated using the least squares methodology and an ARCH(1) specification are compared and it is reported that the larger the magnitude of conditional heteroscedasticity, the larger the difference between least squares and ARCH betas. Bera *et al.* (1988) argue that when the characteristics of variance are considered directly, in other words, conditional heteroscedasticity is modelled, the result is a more realistic estimate of model coefficients, improved efficiency and a reflection of omitted variables in the coefficient estimates (see also Bollerslev & Wooldridge, 1992: 156; Hamilton, 2010). It follows that if variance is assumed to follow an ARCH/GARCH specification, the structure of the variance is reflected in the log-likelihood function and by implication, the parameters of the model. As with the parameters of a multifactor model (the conditional mean in ARCH/GARCH terminology), ω, α_i (in equation 6.24) and β_i (in equation 6.25) are also estimated by maximising the log-likelihood function (Herwartz, 2004: 202).

The link between factor omission and the associated underspecification, the log-likelihood function and the structure of the conditional variance may be established by making a distinction between pure and impure heteroscedasticity. Pure heteroscedasticity persists even when a model is correctly specified and is therefore inherent to the data. Impure heteroscedasticity arises when a relevant factor is omitted (Bucevska, 2011: 631). Including relevant and previously omitted factors in the linear factor model reduces (or eliminates) impure heteroscedasticity and thereby affects the structure of the conditional variance which enters the log-likelihood function in equation (6.23). This, in turn, impacts coefficient estimates. Using the ARCH/GARCH framework permits the level of underspecification, which is now also reflected in the structure of the conditional variance, to be reflected in the coefficient estimates of the conditional mean. This permits an investigation of the impact of underspecification on coefficient estimates across the restricted and unrestricted specifications and the benchmark model even when orthogonal factors are included and excluded from a specification. Accordingly, Armitage and Brzeszczyński (2011: 1533) state that the volatility of returns is a proxy for the impact of information on stock prices and propose that volatility may explain differences in coefficient estimates. In addition, the ARCH/GARCH methodology is appropriate for modelling return series in the presence of non-normality in the form of excess kurtosis and non-linear dependence (Table 6.2. and Table 6.3.) and yields more efficient coefficient estimates relative to the least squares methodology (Elyasiani & Mansur, 1998: 548; Andersen *et al.,* 2003: 48).

This study follows a similar approach to that of Armitage and Brzeszczyński (2011: 1529) in selecting the appropriate ARCH*(p)* or GARCH*(p,q)* specification and the number of ARCH and GARCH parameters.[101] An ARCH(1) model with conditionally normal errors is initially estimated. If there are any remaining ARCH effects, as established by applying the ARCH LM test at lower and higher orders (ARCH(1) and ARCH(5)), and/or the residuals continue to exhibit the presence of non-linear dependence indicative of non-stationary variance, as established by applying Q-statistics at the first and fifth orders ($Q(1)$ and $Q(5)$), an ARCH(2) or GARCH(1,1) model is estimated (Akgiray, 1989: 64).[102] Although not necessarily estimated for each series, the GARCH(1,1) is considered to be the simplest and most robust model of the ARCH/GARCH family of models (Engle, 2001: 166). Bollerslev, Chou and Kroner (1992: 10) state that for most applications, the number of ARCH and GARCH parameters set to *p*=*q*=1 is sufficient. However, if ARCH effects are still present in the residuals and variance is non-stationary, the number of *p* (ARCH) and *q* (GARCH) terms is increased until the residuals are free of ARCH effects and non-linear dependence. This approach is applied to the benchmark model and the conditional variance structure established for each return series in the benchmark model is the initial (starting) structure for the restricted and unrestricted models. As the conditional variance is more likely to reflect impure heteroscedasticity in these specifications, a more complex conditional variance structure may be required to ensure that the residuals are free of ARCH effects and non-linear dependence. Therefore, as in Armitage and Brzeszczyński (2011), increasingly complex ARCH*(p)* or GARCH*(p,q)* specifications are fitted to the conditional variance of the residuals of the restricted and unrestricted models. This permits for a comparison of the conditional variance structures across specifications and also reflects the impact of the inclusion of the residual market factors on coefficient estimates - even if these factors are orthogonal to the macroeconomic factor set. Therefore, it is assumed that information associated with omitted factors is reflected in the (changing) conditional variance structure

[101] The conventional approach to selecting an appropriate ARCH/GARCH model with the associated number of ARCH and/or GARCH parameters for each series is to use model selection criteria such as the Akaike Information Criterion (AIC), Schwarz Information Criterion (SIC) and Bayesian Information Criterion (BIC) (Cornish, 2007: 86-88; Javed & Mantalos, 2013).

[102] Bera *et al.* (1988: 209) show that the impact of ARCH effects on coefficient estimates is dependent upon the numerical magnitude of conditional heteroscedasticity as measured by ω in equation (6.24) and equation (6.25). The lower the magnitude, the lower the difference between least squares and ML coefficient estimates. Therefore, in the absence of ARCH effects, fitting an ARCH(1) model will have no impact on coefficient estimates. In this study, this is confirmed in (unreported) preliminary analysis by comparing least squares and ML coefficient estimates in specifications of the linear factor model. Preliminary analysis also indicates that the GARCH(1,1) model is sufficient for removing ARCH effects in the residuals of specifications that omit factors where an ARCH(1) fails to do so in such specifications.

underlying each residual series. It is further assumed that this information is reflected in the coefficients of the linear factor model. This is because the structure of the conditional variance enters the log-likelihood function (Singh *et al.,* 2010: 59-62; equation (6.23)).

The estimation of ARCH*(p)* and GARCH*(p,q)* specifications requires an assumption about the conditional distribution of the residuals. In estimating the ARCH*(p)* and GARCH*(p,q)* specifications, conditionally normal errors (residuals) are assumed but are treated as an approximation. Nwogugu (2006: 1741) argues that stock prices do not follow any specific distribution and states that any assumed conditional distribution is a “very rough” approximation. This implies that regardless of which distribution is assumed, conditional errors will not be described perfectly. An added benefit of assuming conditionally normal residuals is that the familiar best linear unbiased properties of the estimators (BLUE) may be retained if the sample is sufficiently large. In this study, the ARCH(*p*) and GARCH(*p,q*) models with normally distributed conditional residuals are applied as a simplifying abstraction for the purposes of estimating the benchmark model and the other specifications applied in the investigation of the consequences of underspecification and the ability of the residual market factor to mitigate factor omission. Forecasts of volatility are of interest insofar as relevant for inference and the estimation of the linear factor model specifications considered (Andersen *et al.,* 2003: 46, footnote 8). Nevertheless, a misspecification of the conditional error distribution can confound the effects of underspecification with the effects of a misspecified conditional residual distribution.

A divergence of the normal distribution from the true distribution may result in an increase in the variance of the estimated coefficients and may lead to inconsistent estimates of model parameters (Fan, Qi & Xiu, 2014: 178). To address this, for series with residuals that are not conditionally normally distributed in the benchmark model and the restricted and unrestricted specifications, as established by applying the JB normality test (Varga & Rappai, 2002: 132), quasi-maximum likelihood (QML) estimates with Bollerslev-Wooldridge robust standard errors and covariance are obtained. QML estimation assumes a normal distribution even if this is not the case and derives maximum likelihood estimates from the normal-based likelihood function. In this manner, consistency and asymptotic normality are achieved although ML asymptotic efficiency is sacrificed (Mittelhammer, Judge & Miller, 2000: 247; Fan *et al.,* 2014: 178).

Varga and Rappai (2002: 136) argue that as with heteroscedasticity, the non-normality of the residuals may invalidate certain significance tests. Jarque and Bera (1987: 164) state that the violation of the normality assumption of the residuals may lead to the use of suboptimal estimators, invalid inferences and inaccurate conclusions. Nevertheless, Fiorentini, Calzolari and Panattoni (1996: 416), show that QML estimation under misspecification of the conditional error distribution will result in residual variance that closely approximates the "true" variance and parameters will also be asymptotically consistent (also see Wooldridge, 2013: 815). Furthermore, Mittelhammer *et al.* (2000: 247) argue that the issue of asymptotic efficiency is only relevant if the correct conditional error distribution can be specified. As suggested by Nwogugu (2006), this is not an easy undertaking. Therefore, this suggests that the application of QML estimation permits for reliable and feasible comparisons of results between specifications, even if the correct conditional residual distribution is not fully known.

6.4.3. Treatment Of Confounders: Overspecification And Multicollinearity

Both overspecification and multicollinearity may confound the effects of underspecification on model results and need to be addressed accordingly.

Studenmund (2014: 186-187) states that the inclusion of an irrelevant factor(s), namely overspecification, does not result in coefficient bias if the coefficient associated with the redundant factor is zero but inflates the standard errors and, in turn, results in lower associated test statistics. This resembles the consequences of underspecification (consequences 3) and 4) in Section 5.3.1.). One approach to resolving overspecification is to exclude redundant factors on the basis of their test statistics or *p*-values – a data-driven approach. Another criterion that may be considered in determining whether a factor should be retained is theoretical in character; consideration may be given to the purpose of the specification (Freund, Wilson & Sa, 2006: 238-239).

The theoretical basis of this study is the APT which proposes that the linear factor model is described by multiple factors and exposures to these factors are potentially associated with compensation in the form of risk premia (Section 2.2.). The APT also proposes that multiple return series, whether returns on portfolios or individual stocks, are described by a common factor structure. The purpose of the linear factor model is to describe return behaviour using a set of factors that approximate the underlying common influences in stock returns. In this study, the purpose is to determine whether a specific factor structure, developed in Chapter

7, can sufficiently describe return behaviour by approximating the underlying common factors and whether the residual market factor can resolve underspecification. As such, in investigating the ability of these factors to approximate the factor structure, these factors should be considered jointly as a set in the benchmark model, the restricted and the unrestricted specifications. Freund *et al.* (2006: 289) argue that if the structure of a model is of primary interest, factor selection may be counterproductive as it may omit factors that could provide relevant information. Consequently, given the spirit of the APT, the approach that is followed is to retain the developed factor structure, which comprises the macroeconomic factors and the residual market factors in the respective restricted and the unrestricted specifications. The factors that comprise the factor analytic augmentation in equation (6.20) are treated differently. As the inclusion of these factors in the benchmark model is an econometric correction for omitted factors and interpretation is not of interest, only the factors that are statistically significant in the regression are retained.[103]

The consequences of multicollinearity also partially mimic those of underspecification. Multicollinearity results in inflated standard errors, wider confidence intervals, unstable coefficients and coefficients of implausible magnitudes (Studenmund, 2014: 265-271; Williams *et al.,* 2013: 11). Furthermore, Mela and Kopalle (2002: 664, 672, 674) argue that depending upon the correlation structure, variance estimates may even decrease in the presence of increasing multicollinearity. This confounds the impact of remedial measures for underspecification with those of multicollinearity. Multicollinearity is further compounded by the failure of diagnostic measures to correctly indicate problems associated with collinearity (as opposed to underspecification) and the failure of diagnostics to account for differences between positive and negative correlation structures.[104] Fortunately, the presence of multicollinearity will not impact the assessment of the ability of the residual market factors in the unrestricted specifications and the factor analytic augmentation in the benchmark model to reduce and ideally eliminate underspecification. While multicollinearity may arise from significant correlation within the macroeconomic factor set, the step-wise incorporation of $M\varepsilon_t$ and $I\varepsilon_t$ in equation (6.22) and then (6.19) and the factors in the factor analytic augmentation set in equation (6.20) will not compound any existing multicollinearity as these factors are orthogonal to the macroeconomic factor set and each other by

[103] As a robustness check, significance is cross-referenced using ordinary and rank correlation coefficients.
[104] Mela and Kopalle (2002: 673) show that the variance of estimated parameters is lower in a negatively correlated environment relative to a situation in which explanatory factors are uncorrelated or are positively correlated.

construction (Czaja *et al.,* 2010: 130). Moreover, because the statistical factors are of interest insofar as they are an econometric correction for underspecification and not for the purposes of interpretation, these factors are included selectively only when statistically significant. This is to avoid the artificial inflation of coefficient standard errors (Greene, 2012: 98). Consequently, any changes or improvements in estimation results, such as changes in the conditional variance structure, reductions in standard errors and increases in the number of significant coefficients, can be attributed to the inclusion of the residual market factors in equations (6.22) and (6.19) and the statistical factors in equation (6.20) and not to multicollinearity induced by these factors.

6.4.4. Results And Model Assessment

The investigation of the impact of underspecification begins with an overview of the benchmark model in Chapter 8. The significance and direction of the impact of the factors in the specification are discussed first. This permits a comparison to other similar studies, an interpretation of the results and a confirmation that the estimated relationships meet *a priori* expectations and are theoretically sensible. The mean coefficients, the number of statistically significant positive, negative and total significant instances observed, together with the mean standard errors and (absolute) *z*-scores are reported. Also reported in the abridged results, are the mean least squares coefficients for the benchmark specification and the associated differences between the ML coefficient estimates for the restricted and unrestricted specifications and the least squares coefficient estimates for the benchmark model. This permits a quantification of coefficient bias and facilitates a comparison of changes in bias in Chapter 9 and Chapter 10 following the exclusion and inclusion of factors. The unabridged results are reported in Appendix A. Bera *et al.* (1988: 209) show that the lower the level of conditional heteroscedasticity that can be attributed to impure heteroscedasticity, as reflected in the coefficient on the ARCH term, α_i, in equation (6.24) and (6.25), the smaller the differences between least squares and ML coefficient estimates. Therefore, because ML coefficients will reflect impure heteroscedasticity associated with underspecification, smaller differences should be observed for better specified models as impure heteroscedasticity is reduced. To determine whether the differences in the means of the least squares and ML coefficients are statistically significantly, a paired-sample *t*-test is applied to the series of least squares and ML coefficients for each factor. The non-parametric Wilcoxon matched-pairs signed-rank test is also applied as a confirmatory test with a null hypothesis that the median difference is zero[105] (Chen & Jordan, 1993: 80; Artiach *et al.,* 2010: 39, 42). Differences in results are noted if there are discrepancies between the results of these two tests.

In Chapter 9 and Chapter 10, comparisons of the model intercepts and coefficients across the specifications are also made. The comparison of intercepts is motivated by the findings of Lehmann and Modest (1987: 259), who suggest that the factor structure will impact the magnitude of the intercepts and will therefore have an impact on inferences relating to performance. That intercepts will differ in magnitude is expected in models that are underspecified. This can be attributed to the bias in the intercept terms that follows from the impact of omitted factors now reflected in the intercepts (consequence 2) in Section 5.3.1.). Similarly, it is expected that underspecification will impact the coefficients associated with the macroeconomic factors. This follows from the application of the ARCH/GARCH methodology to model the structure of the conditional variance which will be impacted by factor omission. As in Singh *et al.* (2010 : 59-62), who argue that volatility is a proxy for information, it follows that factor omission represents the omission of information that will be reflected in the conditional variance. This, in turn, will be reflected in model coefficients. It is anticipated that if factor omission translates into a significant coefficient bias, then differences between mean coefficients will be statistically significant. Specifically, the coefficients of the restricted model will differ from those of the benchmark model and the coefficients of the unrestricted models will differ from those of the restricted model if underspecification impacts coefficient estimates. If the conventional residual market factor resolves underspecification, then differences between coefficients of the unrestricted market model and the benchmark model should be insignificant.

To measure the explanatory power, the predictive ability and the ability of the specifications to approximate the true return generating process, the mean, minimum and maximum $\bar{R}^2$ values and the Akaike Information Criterion (AIC) and Schwarz Bayesian Information Criterion (BIC) statistics are reported. The $\bar{R}^2$ is indicative of the proportion of the total variation in returns that is explained by the model (Greene, 2012: 81). The AIC and BIC statistics present two different approaches to model comparison but both are useful in this

[105] The Wilcoxon matched-pairs signed-rank test is applied as a confirmatory test as there may be ambiguity relating to the distribution of the coefficient values.

study. The AIC and BIC statistics can be defined as follows (Javed & Mantalos, 2013: 1920-1921; Spiegelhalter *et al.,* 2014: 1-2):

$$AIC_i = -2\ln L + 2k \tag{6.26}$$

$$BIC_i = -2\ln L + k\ln T \tag{6.27}$$

where in equation (6.26) and (6.27), L is the maximised value of the log-likelihood function, k (in equation (6.26)) is the number of factors in a model and T (in equation (6.27)) is the number of observations. In the respective equations, the $2k$ and $k\ln T$ terms are the respective penalty components. The AIC indicates which model yields the best predictions of actual observed data. In doing so, this measure indicates the level of bias for models estimated using ML estimation (Konishi & Kitagawa, 2008: 60). In contrast, the BIC identifies the specification that best approximates the true return generating process (Spiegelhalter *et al.,* 2014: 1 – 2). The usefulness of these two comparative approaches is immediately evident. The restricted model can be compared to the benchmark model and the unrestricted models can be compared to the restricted model and the benchmark model. This permits for it to establish whether these specifications approximate the performance of the benchmark specification in terms of predictive ability and whether these specifications are comparable approximations of the true return generating process. Unlike the $\bar{R}^2$, the AIC and BIC are not a measure of fit in the classical sense that offers immediate interpretation. Rather, these statistics permit a comparison of a specification to an alternative(s). Comparisons are made across specifications in Chapter 9 and Chapter 10 although these measures are also reported in Chapter 8.

To compare the restricted and unrestricted specifications to each other and the benchmark specification, the changes in the number of significant coefficients are first discussed. To compare specifications on the $\bar{R}^2$, the AIC and the BIC statistics, the approach of Chen and Jordan (1993: 80) is followed. The paired-sample *t*-test is first applied, followed by the Wilcoxon matched-pairs signed-rank test as a confirmatory test (as before) to test the significance of the differences in the $\bar{R}^2$, AIC and BIC values across specifications. If the restricted model approximates the benchmark model, and is therefore adequately specified, differences between these measures should be insignificant. If the inclusion of the residual market factors improves the restricted model, then the differences between these measures for the restricted specification and the unrestricted models should be statistically significant.

If the unrestricted specifications approximate the fully specified benchmark model, then the differences in these measures should not be statistically significant.

It is hypothesised that if the conventional residual market factor is an adequate proxy for omitted factors, then the unrestricted market model should not differ significantly from the benchmark specification. There should be no significant differences between the intercepts and mean coefficients of the benchmark model and the unrestricted market model. The $\bar{R}^2$ and the AIC and the BIC measures for the unrestricted market model should not differ from those of the benchmark model (equation (6.20)). Additionally, any further improvements from the inclusion of a second residual market factor should be marginal.

6.4.5. Model Diagnostics, Robustness And Comparisons Assessment

The results of the diagnostic tests are compared across specifications to investigate whether factor omission (in the restricted and unrestricted models) impacts model diagnostics. The primary comparisons focus on differences across specifications at the individual sector level (for example, the number of significant instances of serial residual correlation). As a secondary form of analysis, consideration is also given to the magnitude of the mean test statistics across specifications. These reflect changes in size, subject to the caveat that all other parameters such as the number of observations and the number of factors remain constant across specifications.[106] (Sullivan & Feinn, 2012).

The first test that is applied is Wald's test of linear restrictions. This is a formal test of the null hypothesis that the $\bar{R}^2$, representative of the overall model fit, is equal to zero. This test determines whether all slope coefficients in a given specification are simultaneously equal to zero and a rejection of the null hypothesis confirms the significance of the multifactor specification (Sadorsky & Henriques, 2001: 204; Studenmund, 2014: 167). The resultant *F*-statistic follows the *F*-distribution with the null hypothesis set out as follows (Blackwell, 2008: 2-3):

$$b_{1k} = b_{2k} = b_{3k} = ... = b_{ik} \tag{6.28}$$

And the alternative hypothesis:

[106] For example, the estimation of the *F*-statistic (equation (6.30)) will also be impacted by the changes in the number of factors in a respective specification whereas *Q*-statistics (equation (6.6)) will reflect a pure size effect if the number of observations remains constant (as it does in this study) (also see Kluve, Schneider, Uhlendorff & Zhao, 2012: 600).

$$b_{1k} \neq b_{2k} \neq b_{3k} = \ldots \neq b_{ik} \tag{6.29}$$

where b_{ik} represents the sensitivity of return series *i* to factor *k,* the factor set in equation (6.21) and also $M\varepsilon_t$ and $I\varepsilon_t$ in equations (6.22) and (6.19) respectively and the factors that comprise the factor analytic augmentation in equation (6.20). The *F*-statistic is defined as (Blackwell, 2008: 4):

$$F = \frac{\left(SSR_r - SSR_{ur}\right)/q}{SSR_{ur}/(n-(k+1))} \tag{6.30}$$

where SSR_r is the sum of the squared residuals of a restricted version of the model that excludes factors and assumes that the null hypothesis is true, SSR_{ur} is the sum of the squared residuals of the unrestricted model, *n* is the number of observations, *k* is the number of explanatory factors and *q* is the number of factor coefficients that are tested. The *a priori* expectation is that the null hypothesis of coefficients jointly equalling zero will be rejected for all specifications and for all series. This is because the macroeconomic factors are chosen for their ability to proxy for the underlying influences in stock returns and therefore should have a significant overall impact in each of the specifications (Section 6.3.2.; Chapter 7). The same may also be said about $M\varepsilon_t$ and $I\varepsilon_t$.

Reported next are the results of the JB test which is applied to the standardised residual series for each specification (Section 6.2.2.). A rejection of the normality assumption indicates that the normal conditional residual distribution is misspecified. For this reason, QML estimation is applied to sectors for which the residual series depart from normality (Section 6.4.2.). Downing and Clark (2010: 403) argue that outliers may be associated with omitted factors, implying that non-normality in the residuals is caused by outliers in return series that are explained by macroeconomic events. It is expected that if underspecification contributes to the non-normality of the residuals, then the number of departures from normality should be lowest for the benchmark specification and highest for the restricted specification. The incorporation of the residual market factors should lower the number of instances of significant departures from normality in the residuals.

Factor omission can induce residual serial correlation. As with heteroscedasticity, serial correlation may be impure and attributable to omitted factors (Mutsune, 2008: 6; Studenmund, 2014: 325). In the presence of residual serial correlation, regression coefficients will be inefficient and conventional tests may be invalid (Granger & Newbold, 1974: 111). Wooldridge (2013: 414) further argues that the nature of the bias in coefficient standard errors is dependent upon whether serial correlation is positive or negative. If serial correlation is positive (negative), then coefficient standard errors will be understated (overstated). This has the potential to result in a misidentification of the linear factor model and the APT relation by overstating (understating) the importance of certain factors (consequence 6) in Section 5.3.1.). To test for serial correlation in the residuals, *Q*-statistics are estimated for first order serial correlation ($Q(1)$), and for the first five serial correlation coefficients ($Q(5)$) (Sadorsky & Henriques, 2001: 206). In the latter case, this test determines whether the first five serial correlation coefficients are jointly statistically significant (equation (6.6)). Adams and Coe (1990: 251) and Claar (2006: 2183) consider the use of alternative specifications to remove impure serial correlation from the residuals. Each of the specifications considered in this study is an alternative specification relative to the others. The unrestricted specifications are alternate reduced form versions of the benchmark model. Similarly, the restricted specification is a reduced form version of the unrestricted and the benchmark models. It follows that a greater number of instances of significant serial correlation should be attributable to specifications that omit factors. The expectation is that the benchmark specification should generally be free of residual serial correlation and any remaining significant residual serial correlation will be pure in nature, unless the functional form of this model is incorrect. If the residual market factors are effective proxies for omitted factors, the number of instances of residual serial correlation should approximate that of the benchmark specification. Finally, it follows that the highest number of instances of statistically significant residual serial correlation will be observed for the restricted model if factor omission introduces impure residual serial correlation. The mean $Q(1)$ and $Q(5)$ statistics and the number of statistically significant instances for each *Q*-statistic are reported in Chapter 8, Chapter 9 and Chapter 10.

A possible cause of residual serial correlation is incorrect functional form. For example, a linear factor model is specified and estimated whereas the correct function form is non-linear (Gujarati & Porter, 2009: 315). This must be excluded as a cause of residual serial correlation so as not to confound incorrect functional form with impure residual serial correlation induced by factor omission. Reinganum (1981: 320) recognises a non-linear functional form as a possible inconsistency in the APT and reports that the APT is unable to explain the firm-size effect. The author attributes this failure of the APT to a number of hypotheses that are tested concurrently and the difficulty in disentangling the hypotheses that are not supported. Notably, Reinganum (1981) argues that the return generating process may not be linear as expounded by the APT. To investigate and exclude non-linearity as a cause of the serial correlation in the residuals of the benchmark model,[107] as opposed to factor omission, a least squares version of the model for sectors that exhibit residual serial correlation is re-estimated. To test for specification error, the Ramsey (1969) regression specification error test (RESET), is applied. This test incorporates polynomials of the fitted factor (returns) in a given specification to test for functional misspecification (Wooldridge, 2013: 306). A failure to reject the null hypothesis of omitted squared factors implies that residual serial correlation is caused by functional form and is not attributable to impure or pure serial correlation. In contrast, in the absence of functional misspecification, any reductions in residual serial correlation can be attributed to the inclusion of relevant factors. Remaining serial correlation will be pure and is an inherent characteristic of the underlying distribution of the residuals and cannot be remedied (Studenmund, 2014: 325).

The econometric methodology and approach outlined in Section 6.4.2. requires that the residuals of the specifications are free of non-linear dependence and ARCH effects for two reasons. First, this study aims to investigate the impact of underspecification on the underlying conditional variance structure, as modelled by the ARCH(*p*) or GARCH(*p,q*) specifications set out in equations (6.24) and (6.25) (discussed in greater detail in Section 6.4.6.). The approach followed is that of Armitage and Brzeszczyński (2011), whereby increasingly complex ARCH(*p*) or GARCH(*p,q*) specifications are estimated for each series until the residual series are free of non-linear dependence and ARCH effects. Second, the presence of heteroscedasticity, which in itself can be impure, will translate into overstated (or understated) standard errors and misleading inferences. Yet, inflated residual variance and the consequent upward bias of the standard errors are also a consequence of underspecification (consequence 3), 4) and 6) in Section 5.3.1.).Therefore, residuals must be free of heteroscedasticity and non-linear dependence so as not to confound the effects of underspecification with those of residual serial correlation. To confirm that residuals are free of non-linear dependence, the approach outlined in Section 6.2.2. is followed. *Q*-statistics are estimated for *squared* residuals for the first serial correlation coefficient, $Q^2(1)$, and up five orders of serial correlation, $Q^2(5)$ (equation 6.6). The ARCH LM test is also

[107] It is hypothesised that if the benchmark model is linear in functional from, then the same may be said about the restricted and unrestricted specifications, which are reduced form models of the benchmark model.

applied to test for first and fifth order ARCH effects (ARCH(1) and ARCH(5); equation (6.8) and equation (6.9)) (Engle, 1982: 999; Sadorsky & Henriques, 2001: 206; Cryer & Chan, 2008: 282). It is anticipated that the residual series obtained from the benchmark, restricted and unrestricted specifications will be free of non-linear dependence and ARCH effects. Mean $Q^2(1)$ and $Q^2(5)$ statistics are reported for tests of joint serial correlation in the squared residual series, and for the ARCH(1) and ARCH(5) LM tests, mean *F*-statistics are reported. The number of significant instances is noted for each specification. It is anticipated that residuals will be free of non-linear dependence and ARCH effects for all specifications.

As a check on the robustness of the benchmark model, the specification for each series is re-estimated using MM and least squares estimation with HAC standard errors (Newey & West, 1987; Andersen *et al.*, 2003: 47-48). MM estimators, developed by Yohai (1987), belong to a class of linear regression estimators that are robust and efficient in the presence of outliers. The technique combines a high breakdown point of 0.5 and high relative asymptotic efficiency[108] to address the limitations of other robust techniques such as S estimation and M estimation (Yohai, 1987: 643). HAC estimation produces standard errors that are robust to serial correlation in the residuals and heteroscedasticity of unknown form (Wooldridge, 2013: 432). Therefore, while HAC estimation confirms the robustness of the results for sectors that exhibit residual serial correlation (if any), MM estimation confirms the robustness of the results if outliers are present in the data (Table 6.2.). Although the results of these regressions are not reported, deviations in results for individual sectors are noted and briefly discussed in Section 8.4. (model diagnostics and robustness). For the restricted and unrestricted versions of the models, discussed in Chapter 9 and Chapter 10, only the sectors that exhibit serial correlation in the residuals are re-estimated using least squares estimation with HAC standard errors. Deviations from ML estimation results are noted to determine whether induced serial correlation impacts model results and inferences (Section 9.4. and Section 10.4.). MM estimation is not applied to the restricted and unrestricted models.

6.4.6. Variance And Conditional Variance

A consequence of underspecification is that residual variance will be biased upwards. Residual variance will partially reflect the squared loadings on omitted factors (Lehmann,

[108] The breakdown point is defined as the maximum fraction of outliers that a sample may contain before the model is impacted (Yohai, 1987: 643). Andersen (2008: 10) reports that MM estimators have an efficiency of approximately 95% relative to the least squares methodology.

1990: 72). Dominguez (1992: 97, 98) argues that misspecification in the form of omitted factors will be reflected in the dispersion of the residuals. This suggests that a comparison of the magnitude of the residual variance across the benchmark specification, the restricted model and the unrestricted models will provide further insight into the impact of underspecification. Importantly, coefficient standard errors are a function of the residual variance (Brooks, 2008: 47). If residual variance is biased upwards as a result of factor omission, coefficient standard errors will also be biased upward. This has the potential to result in an erroneous tendency not to reject the null hypothesis of a coefficient equalling zero (Wooldridge, 2013: 99).

Assuming that as in equation (3.3) (restated), $\sum_{j=1}^{J} b_{ij} f_{jt}$ represents an omitted factor or set of omitted factors and the associated sensitivities:

$$\varepsilon_{it} = +\sum_{j=1}^{J} b_{ij} f_{jt} + \varepsilon_{it}^{*} \tag{3.3}$$

Then $\varepsilon_{it} > \varepsilon_{it}^{*}$ as ε_{it} also reflects omitted factors, f_j s. The ML estimator of variance, assuming a normal distribution, is given by:

$$\sigma_{\varepsilon_i}^2 = \frac{\sum \varepsilon_{it}^2}{n} \tag{6.31}$$

where $\sigma_{\varepsilon_i}^2$ is residual variance and n is the number of observations. This implies that if factors are omitted then:

$$\frac{\sum \varepsilon_{it}^2}{n} > \frac{\sum \varepsilon_{it}^{*2}}{n} \tag{6.32}$$

Accordingly, estimated residual variance will be overstated. This also implies that the standard errors of model coefficients will be overstated, as denoted by:

$$se(b_{ik}) = \frac{\sigma_{\varepsilon_i}^2}{\sum \hat{f}_k^2} > \frac{\sigma_{\varepsilon_i}^{*2}}{\sum \hat{f}_k^2} \tag{6.33}$$

where $se(b_{ik})$ in equation (6.33) is the standard error for the coefficient on factor *k,* f_k, and $\hat{f}_k$ is the difference between an observed individual innovation for factor *k* at time *t* and the mean of factor innovations, $\bar{f}_k$.

The consequences of inflated residual variance stemming from factor omission are therefore two fold. The upward bias will translate into an erroneous failure to not to reject null hypotheses as a result of biased standard errors, lower test statistics and wider confidence intervals (Sykes, 1993; Van Rensburg, 2000: 37; 2002; 91; Studenmund, 2014: 178-200; consequence 4) in Section 5.3.1.). Consequently, the linear factor model may be misidentified. Moreover, if the residual variance or standard deviation is used in tests of the validity of the APT relation, the APT may erroneously be declared invalid. This is because these factors will reflect loadings on omitted factors that may be associated with systematic factors as opposed to purely idiosyncratic ones (Fama & French, 1993; 7-8; Brennan *et al.,* 1998: 366).

It is anticipated that inflated residual variance, $\sigma^2_{\varepsilon_i}$, will impact inferences relating to the overall number of significant factors across sectors in each specification (Section 6.4.4.). To determine whether factor omission in the restricted model impacts residual variance and resultant inferences, mean standard errors are first compared across specifications. The restricted specification should yield the highest mean standard errors if macroeconomic factors fail to adequately characterise the return generating process. If $M\varepsilon_t$ sufficiently resolves underspecification, then standard errors for the unrestricted market model should be comparable to those of the benchmark model. The contribution of $I\varepsilon_t$ should be minor.

As a formal test, the mean residual variance for each model is compared across specifications. The paired-sample *t*-test is applied to the mean residual variance for each specification and the Wilcoxon matched-pairs signed-rank test is applied as a confirmatory test. If factor omission inflates residual variance in the restricted model, then the differences between the mean residual variance for the benchmark and restricted specifications will be statistically significant. If $M\varepsilon_t$ sufficiently reduces residual variance, then the differences between the mean residual variance for the benchmark and unrestricted market model specification should be statistically insignificant.

Differences in the magnitude of residual variance estimates across individual sectors for each specification are also investigated. The Brown-Forsythe test for the equality of variance is applied to the residual series generated for each industrial sector under the benchmark, restricted and unrestricted models (Brown & Forsythe, 1974). The Brown-Forsythe test is a modification of the Levene test of the equality of variance that is robust to departures from normality (Charles, 2010: 146; James & Karoglou, 2010: 481). The Brown-Forsythe test statistic, used in regression modelling applications, is defined as follows (Kutner, Nachtsheim, Neter & Li, 2005: 116-117; 329; 784):

$$BF = \frac{\bar{d}_1 - \bar{d}_2}{s\sqrt{\frac{1}{n_1} - \frac{1}{n_2}}} \tag{6.34}$$

where BF is the Brown-Forysthe test statistic, $\bar{d}_1$ and $\bar{d}_2$ are the sums of absolute deviations of the residuals from their respective medians derived under each specification, n_1 and n_2 are the respective sample sizes and s is the pooled standard deviation.

The Brown-Forsythe test outlined above is applied to test the equality of residual variance estimates across model specifications for each sector. It is postulated that if the residual market factor is an adequate proxy for omitted factors, then the residual variance estimates across the respective specifications (for example, residual variance estimates derived from the benchmark and unrestricted market models for a specific sector) for each sector will be homogeneous. The residual variance estimates derived from the restricted specification are also compared to those derived from the benchmark and the unrestricted models. Significant differences in residual variance estimates will indicate that factor omission inflates residual variance, potentially resulting in misleading inferences.

A consequence of underspecification is that residuals may also exhibit induced impure heteroscedasticity attributable to factor omission. This contrasts with pure heteroscedasticity, which may be present even if the model is correctly specified and is inherent to the data (Bucevska, 2011: 630-631; Gujarati, 2004: 391; Studenmund, 2014: 179). Given the nature of the ARCH/GARCH methodology, both pure and impure heteroscedasticity will impact model parameter estimates by entering the log-likelihood function (equation (6.23)) through the conditional variance, h_{it}, modelled as an ARCH(p) or GARCH(p,q) process (equation (6.24) & (6.25); Engle, 2001: 160). This can explain differences between the respective model coefficients obtained from the benchmark, restricted and unrestricted models even though the residual market factors and the factor analytic augmentation are orthogonal to the macroeconomic factor set and each other. Factor exclusion should translate in a changing conditional variance structure by impacting impure conditional heteroscedasticity (Bera *et al.,* 1988: 209). For this reason, increasingly complex ARCH(*p*) or GARCH(*p,q*) specifications may be required to ensure that the residuals of each specification for each sector are free of non-linear dependence and/or ARCH effects.

Comparing the changing complexity of the ARCH(*p*) and GARCH(*p,q*) specifications across the benchmark, restricted and unrestricted specifications yields an insight into the impact of impure heteroscedasticity on residual variance associated with factor omission and the impact of factor omission on conditional variance in general. Consequently, the use of ARCH(*p*) and GARCH(*p,q*) models allows insight into the consequences of underspecification on the variance and also permits a study of the coefficient bias in the linear factor model, even if orthogonal factors are used. In this study, the ARCH(*p*) and GARCH(*p,q*) structures are compared across specifications. The frequency of each ARCH(*p*) or GARCH(*p,q*) model applied to model the conditional variance structures for each specification of the linear factor model is reported. The number of statistically *F*-statistics is reported for each type of ARCH(*p*) or GARCH(*p,q*) model to confirm the appropriateness of these specifications across sectors. Also reported are the mean values of the intercepts, ω, in the ARCH and GARCH specifications and the respective ARCH and GARCH coefficients, α_i and β_i, for summative reasons. Mean *F*-statistics are not compared for effect size as the number of ARCH(*p*) or GARCH(*p,q*) models applied are likely to differ across sectors and across the linear factor models and therefore, a comparison of mean *F*-statistics will not be meaningful and will be confounded by changing frequencies of the type of ARCH/GARCH models applied. It is anticipated that the conditional variance structures of the benchmark model will be simpler, as represented by a higher frequency of ARCH(*p*) specifications applied, relative to those of the restricted specification which are anticipated to require more frequent GARCH(*p,q*) modelling. If the conventional residual market factor is an adequate proxy for omitted factors, the complexity of the conditional variance structures underlying the unrestricted market model should be comparable to that of the benchmark model. The inclusion of a second residual market factor should not impact conditional variance structures.

Bera *et al.* (1988: 204) state that the ARCH coefficient, α_i, in equations (6.24) and (6.25) quantifies conditional heteroscedasticity, namely heteroscedasticity dependent upon factors in the linear factor model (the conditional mean). This is the coefficient on the squared residual term (ε_{it-p}^2) in these specifications. The authors postulate that the greater the level of conditional heteroscedasticity as measured by the ARCH coefficients, which are dependent upon model specification, the greater the deviation in ML coefficient estimates from least squares coefficient estimates. This suggests that the overall level of underspecification associated with a given specification may be quantified by aggregating the ARCH coefficients of the respective ARCH(*p*) and GARCH(*p,q*) specifications. Comparisons are therefore also undertaken on the basis of the ARCH coefficients to quantify the impact of factor omission on conditional heteroscedasticity across specifications. However, this comparison must be approached with caution. Bera *et al.* (1988) rely upon a single ARCH-type model, the ARCH(1) model and base their inferences upon the magnitude of α_i. In contrast, this study applies the ARCH(*p*) *and* GARCH(*p,q*) specifications. Therefore, it is possible that the level of conditional heteroscedasticity will be reflected in both the α_i and β_i estimates of the GARCH(*p,q*) model. This may make comparisons of the changing levels of conditional heteroscedasticity across specifications solely on the basis of α_i unreliable. The GARCH coefficients, the β_is, will reflect conditional heteroscedasticity and the level of long-run variance captured by the intercept term, ω, in equation (6.24) and equation (6.25) (Dowd, 2002: 316).[109] This is because the GARCH term is the previous forecast of the conditional variance, h_{it-q}, and is dependent upon the constant, ω, and the ARCH term ε_{it-p}^2. Consequently, a portion of conditional heteroscedasticity arising as a result of underspecification will be reflected in the structure of the conditional variance, which may change as a result of a changing linear factor model specification.[110] It follows that comparisons of the structure of conditional variance, as opposed to comparisons of conditional heteroscedasticity, may be more meaningful. Nevertheless, an attempt is made to compare the overall levels of conditional heteroscedasticity in a similar manner as in Bera *et al.* (1988). A measure of the mean level of conditional heteroscedasticity is estimated by

[109] The long-run variance is obtained as follows: $V = \omega / (1-\alpha_i-\beta_i)$. Therefore, ω by itself, is not the long-run variance but rather a representation of the long-run variance; some level to which volatility reverts (Dowd, 2002: 316).

[110] Any change in the conditional mean (the restricted, unrestricted versions or the benchmark model) will also be reflected in the GARCH coefficient.

summing the respective α_i coefficients obtained from each ARCH(p) and GARCH(p,q) and estimating a mean value. The respective mean values are then compared across specifications with the paired-sample t-test applied to determine whether differences are statistically significant. The Wilcoxon matched-pairs singed-rank test is applied as a confirmatory test.

6.4.7. Predictive Ability

The linear factor model represents the return generating process that describes realised returns. Therefore, any proposed specification of the linear factor should yield predictions that resemble realised (actual) returns as closely as possible. To investigate whether factor omission impacts the intertemporal predictive ability (analogously explanatory ability in the present context) of the linear factor model (consequence 5) in Section 5.3.1.), the approach of Chang (1991: 387) is followed. Residuals are estimated for each series and aggregated. The aggregate residual series derived from each specification are reported and then tested against a null hypothesis of mean and median errors equalling zero using the t-test and the Wilcoxon test. It is anticipated that the benchmark specification should produce the lowest residuals and that residuals should not be significantly different from zero. Furthermore, if the residual market factor is an adequate proxy for omitted factors, then the difference between the residuals of the unrestricted market model (and also the unrestricted model) and the benchmark model should be zero. Therefore, the paired sample t-test and the Wilcoxon matched-pairs signed-rank test are applied across specifications to determine whether differences in residual means and medians are statistically significant (Chen & Jordan, 1993: 80; Artiach *et al.,* 2010; 39).

As an additional method of comparing specifications, comparisons are made using Theil's (1966) inequality coefficient in a similar manner as in Chang (1991) and Chen and Jordan (1993). The related decompositions are also considered.[111] Theil's inequality coefficient, the U-statistic, is given by (Brooks, 2008: 254; Brooks & Tsolacos, 2010: 272):

[111] Fair (1984: 264) states that Theil's U statistic may also be used to evaluate *ex-post* forecasts and for comparative purposes across models.

$$U = \frac{\left[\frac{1}{n} \sum_{t=1}^{n} \left(R_{it} - \hat{R}_{it} \right)^2 \right]}{\left[\frac{1}{n} \sum_{t=1}^{n} R_{it}^2 \right]^{1/2} + \left[\frac{1}{n} \sum_{t=1}^{n} \hat{R}_{it}^2 \right]^{1/2}} \tag{6.35}$$

where U is the Theil U statistic, R_{it} and $\hat{R}_{it}$ are observed (actual) returns and predicted returns respectively, and n is the number of observations. The U-statistic is bounded between zero and one and takes on a value of zero if a specification under investigation perfectly predicts the actual observations. Values closer to zero are desirable. A value of one implies that the specification fails to predict observations (predictions are equal to zero) (Bliemel, 1973: 446; Brooks & Tsolacos, 2010: 273). Chang (1991: 387) and Chen and Jordan (1993: 79) compare the U-statistics of two models to determine the superiority of competing specifications. Chen and Jordan (1993) also apply the Wilcoxon matched-pairs signed-rank test to evaluate whether the difference between the U statistics is statistically significant. In this study, the paired-sample t-test is applied to test for significant differences and the Wilcoxon matched-pairs signed-rank test is applied as confirmatory test.[112]

The approach of Theil (1966) has a particular although not unique advantage over other statistical loss functions that is relevant in the present context. The U statistic can be decomposed into bias, variance and covariance proportions (Watson & Teelucksingh, 2002: 139-140; Brooks & Tsolacos, 2010: 272). The bias proportion (U_{BIAS}) indicates the discrepancy that arises between the mean values of the predictions and actual observations. It is a measure of systematic error[113] in the model:

$$U_{BIAS} = \frac{\left(\bar{\hat{R}}_i - \bar{R}_i \right)^2}{\frac{1}{n} \sum_{t=1}^{n} \left(\hat{R}_{it} - R_{it} \right)^2} \tag{6.36}$$

where $\bar{\hat{R}}_i$ and $\bar{R}_i$ are the means of the predicted and actual return values and $\frac{1}{n} \sum_{t=1}^{n} \left(\hat{R}_{it} - R_{it} \right)^2$ is the mean square error (MSE). The closer the bias proportion is to zero, the lower is the level of systematic bias in the specification.

[112] Chang (1991) and Chen and Jordan (1993) use U^2 statistics, a variant of Theil's inequality measure.
[113] Systematic error refers to the extent of consistent over- or underprediction.

The variance proportion, U_{VAR}, indicates the ability of a model to replicate the variance of the actual observations. A high (low) variance proportion indicates that the variability of the actual series is greater (lower) than that predicted by model:

$$U_{VAR} = \frac{\left(\hat{\sigma}_i - \sigma_i\right)^2}{\frac{1}{n}\sum_{t=1}^{n}\left(\hat{R}_{it} - R_{it}\right)^2} \tag{6.37}$$

where $\hat{\sigma}_i$ and σ_i are the predicted and actual standard deviation values for series *i*, respectively. A lower variance proportion bias is desirable, implying that a specification more closely replicates the variance of the actual observations (Kacapyr, 2014: 162). Finally, the covariance proportion, U_{COV}, measures the extent to which prediction errors are attributable to unsystematic errors or residual components. These are errors which remain after deviations from actual values have been accounted for.:

$$U_{COV} = \frac{2\hat{\sigma}_i\sigma_i\left(1 - \rho(\hat{R}_{it}R_{it})\right)}{\frac{1}{n}\sum_{t=1}^{n}\left(\hat{R}_{it} - R_{it}\right)^2} \tag{6.38}$$

where $\hat{\sigma}_i$ and σ_i are the predicted and actual standard deviation for series *i*, respectively and $\rho(\hat{R}_{it}R_{it})$ is the correlation between predictions and actual observations. A large covariance proportion, ideally equal to unity, implies that most of the prediction errors are attributable to the random nature of the phenomena that is being predicted and is unrelated to the ability of the model to replicate the mean or the variance of the actual return series (Elkhafif, 1996: 97; Brooks & Tsolacos, 2010: 272; Kacapyr, 2014: 162).

Watson and Teelucksingh (2002: 140) state that a good model should have a bias proportion of close to zero (below 0.1 or 0.2), a small or non-existent variance proportion and a co-variance proportion that is close to unity. The authors further state that the Theil decomposition provides the most comprehensive analysis of model simulation and favour its use over other approaches. As with the *U*-statistic, comparisons across the bias, variance and covariance proportions are made across specifications using significance tests.

6.4.8. Factor Omission

The comparisons of results across models (Section 6.4.4.), model diagnostics and robustness (Section 6.4.5.), residual variance and the conditional variance (Section 6.4.6.)

and predictive ability (Section 6.4.7.) implicitly (at this stage of the study) assume that the restricted model (equation (6.21)) is underspecified and that the unrestricted specifications (equation (6.19) and equation (6.22)) potentially approximate the benchmark model and are adequately specified if the residual market factor or two residual market factors are proxies for omitted factors. The assumption that the restricted model is (potentially) underspecified and that the unrestricted models are (potentially) adequately specified must be confirmed and is tested using two approaches.

The first approach relies upon the likelihood ratio (LR) test. The LR test is applied to each specification, with the exception of the benchmark model, to determine whether each specification omits factors that are incorporated into expanded form alternatives. The LR test statistic follows a chi-squared (χ^2) distribution and the LR test statistic is estimated as follows (Azeez & Yonezawa, 2006: 575; Asteriou & Hall, 2016; 79):

$$LR = -2(L_R - L_U) \tag{6.39}$$

where L_R and L_U are the maximised values of the log-likelihood function (equation (6.23)) for a restricted log-likelihood function that excludes factors and an unrestricted log-likelihood function respectively. It follows that if the restricted log-likelihood function is appropriate, then this function will not differ from an unrestricted log-likelihood function, which incorporates an omitted factor or a set of factors. If this is the case, then no relevant factors have been omitted and a model is not underspecified (Gujarati, 2004: 295; Asteriou & Hall, 2016; 79).

The restricted model is tested first against the null hypothesis that $M\varepsilon_t$, $I\varepsilon_t$ and finally $\sum_{j=1}^{J} b_{ij} f_{jt}$ (the *full* factor analytic augmentation) separately represent insignificant factors. The same procedure is followed for the unrestricted market model ($I\varepsilon_t$ and $\sum_{j=1}^{J} b_{ij} f_{jt}$ are the hypothesised omitted factors) and for the unrestricted model (the factor analytic augmentation, $\sum_{j=1}^{J} b_{ij} f_{jt}$, is the hypothesised omitted factor set). A rejection of the null hypothesis confirms that a specification is underspecified – relevant factors are omitted.

The second approach is based on that of Van Rensburg (1995: 63), who factor analyses the residual correlation matrix of a linear factor model to derive residual common factors. It also partially follows the work of Meyers (1973). In this analysis, the factor analytic approach is identical to that outlined in Section 6.3.1. The scree and MAP tests are applied to gain insight into the structure of the pairwise residual correlation matrices of the respective specifications considered, including the benchmark model that already incorporates a factor analytic augmentation derived from the residuals of the unrestricted model (equation (6.19)). It is expected that if there are no omitted factors, then no common factors will be extracted from the respective residual correlation matrices. In line with this reasoning, Meyers (1973: 698) argues that if a (linear factor-type) model is valid, then no factors should be reflected in the residual dependence structure.

If the scree and MAP tests point towards the existence of common factors, then this implies that the macroeconomic factor set and the respective residual market factors potentially fail to account for all common influences. This will pose a challenge to the validity of a given specification. To ensure that any extracted factors are not pseudofactors, namely factors that explain common variance for a limited number of series, the resultant factor loadings are analysed. It is possible that any extracted common factors are attributable to strong interdependence between a limited number of series and are not the result of widespread pairwise residual correlation. Consideration is also given to the resultant communalities, which may indicate that the extracted factors are trivial in which case the challenge to the validity of a model is unwarranted.

Pseudofactors can also be defined as factors that are important to specific time periods (Kryzanowski & To, 1983: 37; 42; Connor, 1995: 44). Meyers (1973: 705) states that their existence will not invalidate a model arguing that if the remaining factors represent transitory statistical artefacts, then the validity of the model need not be interrogated. To determine whether the relevance of any of the extracted factors is limited to specific time periods and whether such factors are transitory in nature, the residual correlation matrix derived from each specification is also factor analysed over two subperiods on the basis of the MAP test. Each subperiod comprises half the sample period; January 2001 (2001M01) to December 2008 (2008M12) and January 2009 (2009M01) to December 2016 (2016M12). It may be that factors extracted from the residual matrix spanning the entire sample period arise from high residual interdependence that is confined to a single subperiod. Such factors are transient. If this is the case, a specification will continue to adequately represent the underlying return generating process.

Finally, factor analysis conducted on the residuals has a further function. The measured communalities in the residuals will indirectly quantify the ability of a specification of the linear factor model to account for co-movement in returns (Yong & Pearce, 2013: 82). Communalities will reflect how much co-movement remains in the residuals after accounting for pervasive influence using the macroeconomic factors and the residual market factors (Elton *et al.,* 2014: 157). If macroeconomic factors, the residual market factor and the second residual market factor are adequate proxies for the pervasive influences in returns, the inclusion of these factors should result in reduced mean communalities relative to the restricted model. If these factors are adequate proxies for the pervasive influences in returns, then communalities should approach zero. All that will remain will be the unique variance as measured by uniqueness, approaching a value of one. This can be interpreted as sector-specific variance and inherent randomness (Walker & Madden, 2008: 326).

6.4.9. The Residual Correlation Matrix

The preceding section outlines methodologies that are applied to investigate whether any systematic influences are (still) reflected in the residuals of the linear factor models considered. Central to the study of the linear factor model and potential factor omission is the structure of the resultant correlation matrix. A key assumption underlying the linear factor model is that the residuals are uncorrelated (Section 2.2.; equation (2.2)). A complete absence of interdependence between residual series and congruence with the diagonality assumption, implies the absence of omitted factors and underspecification, and is desirable (Elton & Gruber, 1988: 31; Van Rensburg, 2002: 97; Elton *et al.,* 2014: 157).

The analysis of the respective pairwise residual correlation matrices of the benchmark, restricted and unrestricted specifications begins with less formal descriptive comparisons of the correlation coefficients in the full residual correlation matrices (King, 1966: 157-158; Bilson *et al.,* 2001: 410). If specifications yield adequate descriptions of the return generating process, it is expected that most residual correlation coefficients will be centred around a mean of zero. This should be evident from the abovementioned summaries of the residual matrices for the respective specifications and any reductions in co-movement will be reflected in these summary measures. Comparisons are therefore made across the respective specifications on the basis of the histograms, frequency tables, means and reported minimum and maximum values that summarise the residual correlation matrices of the respective specifications. As an informal and limited test of the magnitude and comparison of residual correlation (see discussion and criticism of this approach below in the outline of the Jennrich (1970) test that follows), the aggregated mean residual correlations from each matrix are tested against a null hypothesis of a mean and median of zero and for statistically significant differences between each other using the paired-sample *t*-test and the Wilcoxon matched-pairs signed-rank test (Eichholtz, 1996: 61).

The next method of analysis is a visual inspection of the residual correlation matrices themselves. This will reveal the extent of instances of statistically significant residual correlation that remain in each resultant matrix. Specifications that do not suffer from factor omission, or are characterised by lower levels of underspecification, should reflect fewer instances of significant residual correlation. It is possible that the residual correlation matrix will reflect economic sector-specific factors that will cause positive residual correlation between industrial sectors, within a specific economic sector. In such instances, a visual inspection should reveal intra-sector positive residual correlation that does not extend beyond the respective economic sectors, if present. Such correlation will not invalidate the linear factor model as these factors, by definition, are diversifiable and have a limited impact (Beenstock & Chan, 1986: 129). Finally, any significant negative correlation coefficients *outside* of the economic sector submatrices will be attributable to between-industry factors (King, 1966: 153). For a well-specified linear factor model, the resultant correlation matrix should reflect no instances of significant residual correlation. If these are present, these should be limited and sporadic in nature. The opposite is true for an underspecified model.

A more appropriate test for comparisons of the structure of the residual correlation matrices relative to a simple *t*-test or Wilcoxon test is the Jennrich (1970) test. By testing the normalized difference between two matrices, this test establishes the equality or lack thereof of two correlation matrices. Unlike the simpler comparisons suggested above, which take into account only the magnitude of correlation and the number of pairwise residual correlations, this test also considers the number of observations from which correlations are estimated. The Jennrich χ^2 statistic follows an asymptotic chi-squared distribution and is estimated as follows (Eichholtz, 1996: 61; Jondeau, Poon & Rockinger, 2007: 30):

$$\chi^2 = 0.5tr(Z^2) - \frac{diag}{(Z)S^{-1}diag(Z)} \tag{6.40}$$

where

$$Z = c^{1/2}R^{-1}(R_{-1} - R_2) \tag{6.41}$$

$$R = (n_1R_1 + n_2R_2)/(n_1 + n_2) \tag{6.42}$$

$$c = n_1n_2(n_1 + n_2) \tag{6.43}$$

and where R_1 and R_2 are the correlation matrices to be compared, n_1 and n_2 are the number of observations upon which the matrices are based. *Diag* denotes the diagonal of a square matrix in column form and *tr* denotes the trace of a second matrix. The Jennrich test statistic has *p(p - 1)/2* degrees of freedom with *p* denoting the dimension of the matrix (in the present case, this will equal 325 degrees of freedom). Testing the equality of matrices allows for a comparison of the residual correlation matrices across specifications and for testing the resultant residual correlation matrices against a null of a diagonal matrix (McElroy & Burmeister, 1988: 41). A diagonal matrix for which the values on the diagonal are one whereas the off-diagonal elements are zero is an identity matrix (Serre, 2002: 5). In this study, this matrix (the identity matrix) is denoted by I_{26}, where 26 represents the dimensions of the matrix.

Testing the null hypothesis that a residual correlation matrix for a given specification is equal to I_{26} is a test of the diagonality assumption, namely the assumption that for a given linear factor model, all off-diagonal entries conform to the assumption of uncorrelated residuals, $E(\varepsilon_{it}, \varepsilon_{it}) = 0$, as denoted by equation (2.2). Consequently, the residual correlation matrices derived from the benchmark model (B_{26}), the restricted model (R_{26}), the unrestricted market model (M_{26}) and the unrestricted model (U_{26}) are tested against the null hypothesis that they are equal to the identity matrix. The residual correlation matrices are also compared to each other to see whether the exclusion of factors (as in the restricted model) and the inclusion of factors (as in the unrestricted specifications) impacts the structure of these matrices. Therefore R_{26} and B_{26} are tested for equality and M_{26} and U_{26} are tested for equality with R_{26} and B_{26} respectively. If the residual market factors are adequate proxies for omitted factors, then B_{26}, M_{26} and U_{26} should not differ significantly. If the residual market factors are proxies for omitted factors that would otherwise be relegated to the residuals, then M_{26} and U_{26} will significantly differ from R_{26}. Finally, each matrix is tested for equality with the correlation matrix of the actual return series, A_{26}. If a model performs poorly, then the matrices shall be equal because much of the correlation structure of the actual return series will also be reflected in the residual correlation matrices.

6.5. CHAPTER SUMMARY AND CONCLUSION

The methodology set out in this chapter is motivated by the theory developed in Chapters 2, 3, 4 and 5. Section 6.2. introduces the return and macroeconomic data that are used in this study. The analysis of the return data indicates that returns depart from normality and exhibit non-linear dependence and ARCH effects at various orders. This points towards the need to use a methodology that can incorporate these characteristics. The initial macroeconomic factor set comprises factors representative of eight categories, namely real activity, prices, cyclical indicators, exchange rates, monetary factors, commodities, interest rates and trade (Table 6.4.). Section 6.2.3. shows that the autoregressive time series methodology employed to generate unexpected components is effective for generating zero-mean innovation series from the macroeconomic factor series. Section 6.3. sets out the methodology employed in exploring the factor structure of the return series in the sample. The aim of this explorative approach is to identify the pervasive influences in stock returns and to identify macroeconomic factors that are proxies for these influences. Also, this approach quantifies the ability of these factors, together with the residual market factors, to proxy for the pervasive influences in returns.

Section 6.4. sets out the proposed specifications of the benchmark, restricted and unrestricted specifications and the econometric methodologies employed in estimating these models. These specifications are central to the comparative research design that is followed in this study. The approach followed aims to investigate the impact of factor omission and the ability of the residual market factors to resolve the consequences of factor omission. The restricted model, which comprises only macroeconomic factors and omits the residual market factors and the factor analytic augmentation that also comprise the benchmark model, is to be compared to the benchmark model. Similarly, the unrestricted models are to be compared to the benchmark model and also the restricted model (Section 6.4.1.). The intention is to determine whether the restricted and unrestricted specifications differ from each other and from the benchmark model according to factor significance, explanatory power, estimated parameters, model diagnostics, residual variance and variance structure and predictive ability. It is hypothesised that underspecification will impact these aspects (Chapter 5).

The methodology applied to confirm factor omission explicitly, as opposed to investigating the consequences of perceived underspecification attributable to the omission of factors, is outlined in Section 6.4.8. and Section 6.4.9. Importantly, Section 6.4.9. outlines the approach employed in analysing the structure of the resultant correlation matrices associated with each specification. The intention is to establish whether co-movement, which is associated with omitted factors is still present after the residual market factors have been incorporated in the unrestricted models. It is hypothesised that if the residual market factor or two residual market factors are an adequate proxy for omitted factors, the unrestricted market model will be comparable to the benchmark model across the numerous aspects outlined in Section 6.4.

Chapter 7 investigates the factor structure of the return series in the sample and identifies the factors that proxy for pervasive influences in South African stock returns.

CHAPTER 7

THE FACTOR STRUCTURE OF THE SOUTH AFRICAN STOCK MARKET

7.1. INTRODUCTION

This chapter explores the factor structure underlying the South African stock market with the intention of identifying the macroeconomic factors that proxy for the pervasive influences in South African stock returns. This is done with the aim of specifying a factor structure that characterises the return generating process of the South African stock market. The factor structure developed in this chapter is employed in Chapter 8, Chapter 9 and Chapter 10 to investigate the consequences of underspecification and the role of the residual market factor.

As in prior work on the APT (Section 2.3.), the exploration of the factor structure begins with a discussion of the number of factors in South African stock returns. As motivated by Connor (1995: 42), a factor analytic approach is followed to identify the number of factors representative of pervasive influences in returns. The subsequent derivation of factor scores can be used to confirm which factors are proxies for influences in returns. This enables a further screening and confirmation of some of the macroeconomic factors identified through a broad preliminary screening based upon the dividend discount model (Table 6.4.). Relevant factors can be confirmed by relating the qualifying macroeconomic factors to the factor scores derived using factor analysis (Section 6.3.). Importantly, this permits for it to be determined how well the macroeconomic factors and also the residual market factors approximate the common influences in South African stock returns. Analysis of this sort also provides preliminary insight into potential underspecification in the models that are estimated in the subsequent chapters, given that macroeconomic factors may not be sufficiently exhaustive proxies for the common factors in returns.

Section 7.2. proceeds by exploring the correlation structure of returns and the number of factors in South African stock returns. The ability of pervasive influences, as measured by derived factor scores, to explain returns is also considered. Section 7.3. undertakes a factor-return correlation analysis with the intention of identifying factors that have a systematic impact on returns. The ability of the identified macroeconomic factors to proxy for common underlying factors in returns is confirmed in Section 7.4. Also considered is the ability of the macroeconomic factors and the residual market factors to approximate the factor structure of South African stock returns. Section 7.5. summarises and concludes this chapter.

7.2. THE FACTOR STRUCTURE OF THE SOUTH AFRICAN STOCK MARKET

The histogram of ordinary correlation coefficients for the return series in the sample is presented in Figure 7.1. and Table 7.1. summarises the distribution of the estimated correlation coefficients.

Figure 7.1: Histogram Of Return Correlation Coefficients

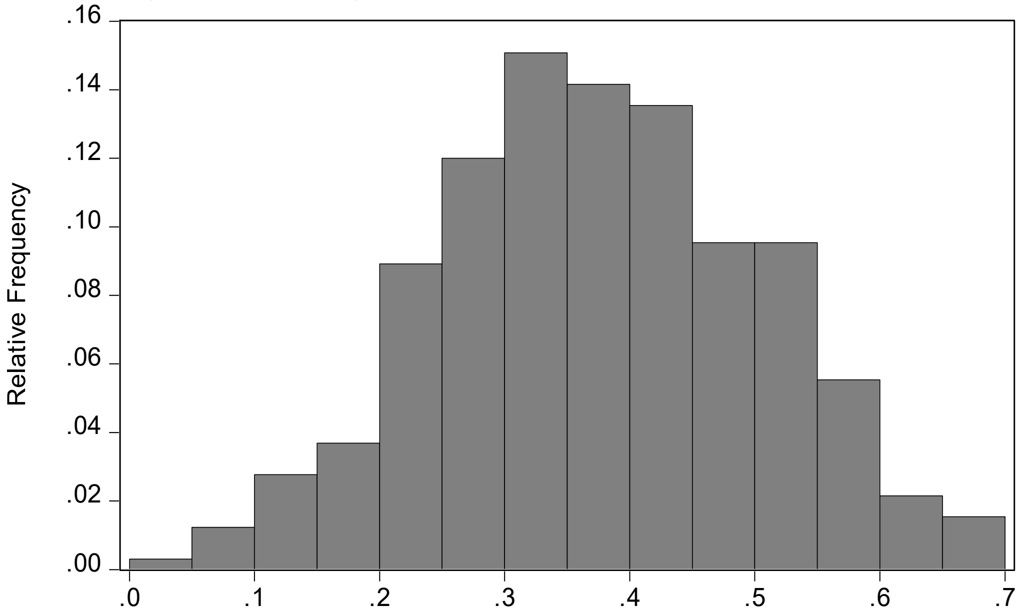


Table 7.1. Summary Of The Distribution Of Return Correlation Coefficients

Bin	Frequency	Relative Frequency	Cumulative Frequency
$-0.1 < \rho_{ij} \leq 0.0$	0	0.000%	0.000%
$0.0 < \rho_{ij} \leq 0.1$	5	1.538%	1.538%
$0.1 < \rho_{ij} \leq 0.2$	21	6.462%	8.000%
$0.2 < \rho_{ij} \leq 0.3$	68	20.923%	28.923%
$0.3 < \rho_{ij} \leq 0.4$	95	29.231%	58.154%
$0.4 < \rho_{ij} \leq 0.5$	75	23.077%	81.231%
$0.5 < \rho_{ij} \leq 0.6$	49	15.077%	96.308%
$0.6 < \rho_{ij} \leq 0.7$	12	3.692%	100.000%
$0.7 < \rho_{ij} \leq 0.8$	0	0.000%	100.000%
Total	**325**	**100%**	**100%**
Mean	0.375***		
Minimum	0.048		
Maximum	0.673		

Notes:

The asterisks, ***, ** and *, indicate statistical significance at the respective 1%, 5% and 10% levels of significance. The *t*-test is applied to test the hypothesis that the mean of correlation coefficients does not differ significantly from zero. The Wilcoxon matched-pairs signed-rank test is applied as a confirmatory test and the superscript "W" indicates that the Wilcoxon matched-pairs signed-rank test contradicts the results of the paired-sample *t*-test. Bin represents ranges of correlation coefficients and Frequency reports the number of correlation coefficients that fall within each range. Relative Frequency is the percentage of correlation coefficients that fall within the respective ranges. Cumulative Frequency is the running total of all previous relative frequencies in percentage terms. Mean is the mean value of the correlation coefficients in the correlation matrix and the Minimum and Maximum are the lowest and largest correlation coefficients observed.

As evident from Figure 7.1, all correlation coefficients are positive. Over half (52.308%) of the off-diagonal correlation coefficients are greater than 0.3 and less than or equal to 0.5 in magnitude. This corresponds to a total of 170 out of 325 correlation coefficients that fall within this range. Only eight coefficients are less than 0.12 in magnitude. Coefficients below this this (absolute) magnitude are generally statistically insignificant. This summary indicates that the correlation structure of returns is overwhelmingly characterised by positive and statistically significant residual correlation. The statistically significant mean level of correlation of 0.375 attests to the sizeable magnitude of the correlation coefficients. Such a result is expected if returns are impacted by the same systematic influences and respond to common influences. Furthermore, correlation coefficients range between 0.048 (minimum) and 0.673 (maximum).

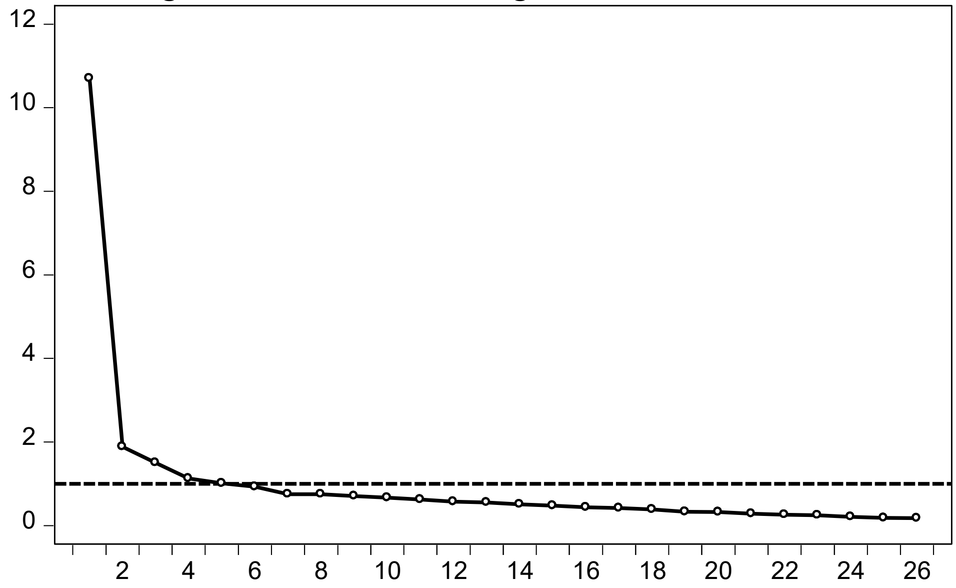
The return correlation matrix is reproduced in Table 7.2. The dashed lines (---) represent economic sector classifications within which the industrial sectors listed in Table 6.1. fall. It is again evident that the return correlation matrix reflects widespread positive interdependence (shaded pairwise correlation coefficients are statistically significant at the 10% level of significance), with very few statistically insignificant instances (as noted above).

Table 7.2: Correlation Matrix Of Returns

	J135	J173	J175	J177	J235	J272	J273	J275	J277	J279	J335	J353	J357	J453	J457	J533	J537	J555	J575	J653	J835	J853	J857	J877	J898	J953
J135	1.000																									
J173	0.367	1.000																								
J175	0.288	0.226	1.000																							
J177	0.334	0.450	0.586	1.000																						
J235	0.492	0.276	0.409	0.281	1.000																					
J272	0.412	0.345	0.234	0.311	0.479	1.000																				
J273	0.544	0.359	0.346	0.339	0.598	0.492	1.000																			
J275	0.461	0.394	0.364	0.396	0.432	0.334	0.596	1.000																		
J277	0.557	0.355	0.289	0.353	0.487	0.538	0.526	0.528	1.000																	
J279	0.462	0.383	0.216	0.225	0.525	0.585	0.513	0.343	0.581	1.000																
J335	0.427	0.284	0.260	0.241	0.224	0.274	0.344	0.352	0.375	0.301	1.000															
J353	0.310	0.422	0.113	0.312	0.124	0.427	0.321	0.271	0.291	0.250	0.229	1.000														
J357	0.428	0.268	0.270	0.318	0.407	0.574	0.525	0.376	0.588	0.513	0.320	0.367	1.000													
J453	0.432	0.286	0.151	0.234	0.378	0.493	0.417	0.326	0.524	0.407	0.242	0.340	0.461	1.000												
J457	0.320	0.123	0.200	0.112	0.312	0.304	0.385	0.259	0.354	0.351	0.262	0.170	0.392	0.418	1.000											
J533	0.286	0.088	0.129	0.048	0.347	0.534	0.292	0.188	0.471	0.416	0.170	0.171	0.491	0.383	0.347	1.000										
J537	0.412	0.189	0.161	0.126	0.463	0.640	0.534	0.372	0.658	0.606	0.326	0.216	0.619	0.507	0.521	0.640	1.000									
J555	0.440	0.293	0.088	0.243	0.306	0.560	0.394	0.301	0.548	0.526	0.274	0.375	0.433	0.421	0.322	0.335	0.533	1.000								
J575	0.480	0.312	0.216	0.230	0.474	0.456	0.443	0.468	0.579	0.537	0.311	0.328	0.491	0.452	0.324	0.282	0.487	0.431	1.000							
J653	0.293	0.121	0.080	0.229	0.183	0.314	0.283	0.193	0.339	0.319	0.205	0.065	0.302	0.174	0.272	0.219	0.364	0.408	0.271	1.000						
J835	0.354	0.233	0.256	0.222	0.414	0.618	0.466	0.362	0.542	0.506	0.208	0.223	0.505	0.396	0.387	0.457	0.669	0.450	0.453	0.319	1.000					
J853	0.359	0.183	0.291	0.241	0.374	0.420	0.420	0.320	0.503	0.337	0.249	0.254	0.496	0.442	0.303	0.383	0.430	0.270	0.384	0.222	0.403	1.000				
J857	0.501	0.436	0.236	0.323	0.365	0.569	0.494	0.471	0.560	0.515	0.328	0.412	0.545	0.469	0.402	0.403	0.577	0.538	0.479	0.325	0.665	0.449	1.000			
J877	0.495	0.335	0.149	0.287	0.418	0.612	0.564	0.355	0.516	0.558	0.286	0.413	0.530	0.343	0.359	0.446	0.634	0.574	0.494	0.384	0.660	0.341	0.673	1.000		
J898	0.412	0.387	0.111	0.308	0.295	0.466	0.363	0.227	0.412	0.432	0.298	0.525	0.419	0.327	0.255	0.299	0.393	0.462	0.399	0.273	0.350	0.344	0.530	0.508	1.000	
J953	0.329	0.340	0.239	0.397	0.244	0.417	0.399	0.277	0.423	0.387	0.288	0.295	0.377	0.277	0.290	0.232	0.367	0.540	0.403	0.397	0.425	0.381	0.571	0.556	0.436	1.000

The results of the factor analysis are discussed below and Figure 7.2. reports the results of the scree test.

Figure 7.2: Scree Plot Of Eigenvalues For Returns



The distinct flexion point indicates that two common factors characterise the structure of South African stock returns. Earlier work on the JSE also supports a two common factor structure for South African stock returns (Barr, 1990; Van Rensburg & Slaney, 1997). Panel A of Table 7.3. reports the eigenvalues and the percentage of variance accounted for by the first 10 factors (Hughes, 1984: 207):

Table 7.3: Summary Of Factor Analysis Of Returns

Panel A: Variance In Returns Accounted For By Each Factor			
Number	**Eigenvalue**	**Proportion**	**Cumulative Proportion**
1	10.704	0.412	0.412
2	1.884	0.072	0.484
3	1.505	0.058	0.542
4	1.123	0.043	0.585
5	1.008	0.039	0.624
6	0.925	0.036	0.660
7	0.752	0.029	0.689
8	0.751	0.029	0.717
9	0.706	0.027	0.745
10	0.669	0.026	0.770
Panel B: MAP Test Results			
Factor(s) Extracted	**Mean Communality**	**Mean Uniqueness**	
3	0.484	0.516	

Notes:
In Panel A, Number is the *n*th extracted factor and Eigenvalue is eigenvalue corresponding to the *n*th factor. Proportion refers to the amount of shared variance explained by the *n*th extracted factor. Cumulative Proportion is the total proportion of variance explained up to the *n*th extracted factor. In Panel B, Mean Communality is the mean proportion of common variance explained across return series by common factors extracted on the basis of the MAP test. Mean Uniqueness is the mean proportion of variance across return series attributable to the return series themselves and not systematic factors.

The first two factors account for almost 50% of the common variation in the returns; the first factor accounts for 41.2% and the second factor accounts for 7.2% of variation. However, the third factor contributes an additional 5.8%. The fourth and fifth factors contribute an additional 4.3% and 3.9% respectively, suggesting that more than two factors may be required to capture the common variation in South African stock returns. An examination of the eigenvalues indicates that the first five eigenvalues are above one, the sixth is close to one (0.925). Although criticised in the literature, Kaiser's (1960) *K1* method proposes that factors with eigenvalues greater than one should be retained for interpretation (Ledesma & Valero-Mora, 2007: 2). The proportion of variation explained continues to decline and is 2.6% for the 10th factor.

In contrast to the results of the scree test, the results of the MAP test indicate that a three factor structure describes the return generating process underlying the South African stock market (Panel B of Table 7.1.). Results indicate a mean communality of 48.4% for a three factor structure suggesting that almost half of the common variation in South African stock returns is accounted for by these three common factors. The first factor accounts for 39.1% of the variation in returns, the second factor accounts for 5.5% and the third factor accounts for 3.8%.

Although the results of the scree and MAP tests are somewhat in conflict and prior findings propose a two factor structure for the South African stock market (specifically that of Barr, 1990)), a three factor structure is chosen to represent the common factors in South African stock returns (Van Rensburg & Slaney, 1997; Van Rensburg, 1998).[114] This is partially motivated by Middleton and Satchell's (2001: 506) argument that if underspecification is to be avoided and there is uncertainty about the number of factors in the model, the principal of parsimony is inappropriate. Following this line of reasoning, it is very likely that more than three macroeconomic factors are required to describe the linear factor model accurately by acting as proxies for influences in the South African stock market.

The extracted factors undergo varimax rotation and each factor score series is retained for use in confirmatory analysis and the identification of proxy macroeconomic factors in the discussion that follows in Section 7.4.

[114] Van Rensburg and Slaney (1997), as cited in Van Rensburg (1998: 22, 32), suggest there are at least two but no more than three factors in South African stock returns. The authors state that it is questionable whether the consideration of more than three factors would contribute meaningfully to the analysis or whether additional factors would be economically interpretable.

7.3. FACTOR-RETURN CORRELATION ANALYSIS AND FACTOR SELECTION

The results of the factor-return correlation analysis are reported in Table 7.4. below.

Table 7.4: Factor-Return Correlation Analysis

Factor	Notation	ρ_P Summary			Sig	ρ_P JSE	ρ_S JSE
		$\rho_P > 0$	$\rho_P = 0$	$\rho_P < 0$			
Panel A: Real Activity							
Manufacturing Sales	MFS_{t-1}	7	19		7	0.201***	0.129*
Wholesale Trade sales	WHL_{t-1}	12	14	0	12	0.141*	0.140*
Retail trade Sales	RET_{t-3}	10	16	0	10	0.084	-0.109
New Vehicle Sales	VEH_{t-1}	12	14	0	12	0.182**	0.136*
Total Mining Prod.	MIP_t	4	22	0	4	0.069	0.065
Building Plans Passed	BP_{t-1}	18	8	0	18	0.226***	0.179**
Buildings Completed	BC_t	0	23	3	3	-0.088	-0.059
Employment	EMP_t	0	24	2	2	0.036	0.046
Panel B: Prices							
Consumer Price Inflation	CPI_t	1	20	5	6	-0.053	-0.012
Inflation Expectations	BAR_{t-1}	0	18	8	8	-0.031	-0.005
Prod. Price Prices	PPI_t	0	18	8	8	-0.062	-0.043
Input Prices	INP_t	0	14	12	12	-0.016	-0.048
Panel C: Cyclical Indicators							
Inventories	INV_{t-1}	0	20	6	6	-0.160**	-159**
Leading Indicator	$LEAD_{t-1}$	15	11	0	15	0.279***	0.187***
Coincident Indicator	$COINC_{t-2}$	8	18	0	8	-0.175**	0.155**
Lagging Indicator	LAG_t	0	20	6	6	-0.044	0.003
House Prices	HSE_t	3	22	1	4	0.069	0.086
Business Activity	BUS_t	17	9	0	17	0.116	0.118*
Panel D: Exchange Rates							
Rand-Dollar Ex. Rate	USD_t	0	7	19	19	-0.114	-0.148**
Rand-Euro Ex. Rate	EUR_t	0	9	17	17	0.041	0.018
Rand-Pound Ex. Rate	GBP_t	0	8	18	18	-0.060	-0.093
Nominal Effective Ex. Rate	NEX_t	18	8	0	18	0.030	0.067
Real Effective Ex. Rate	REX_t	17	9	0	17	-0.017	0.028
Panel E: Monetary Factors							
M0 Monetary Aggregate	$M0_{t-2}$	9	17	9	9	0.144**	0.177**
M1A Monetary Aggregate	$M1A_{t-1}$	10	16	0	10	0.169**	0.183**
M1 Monetary Aggregate	$M1_{t-1}$	9	17	0	0	0.161**	0.104
M2 Monetary Aggregate	$M2_{t-2}$	1	22	3	4	-0.049	-0.058
M3 Monetary Aggregate	$M3_{t-2}$	0	16	10	10	-0.188***	-0.143**
Excess M3 Supply Growth	$M3E_{t-1}$	0	20	6	6	-0.246***	-0.221***
Total Credit Extension	TCR_{t-1}	4	22	0	4	0.178**	0.139*
Private Credit Extension	PVC_t	0	24	2	2	-0.080	0.003
Gold Reserves	GFR_t	2	6	18	20	0.091	0.042
Foreign Reserves (US$)	RES_t	11	15	0	11	0.190***	0.215***
Foreign Reserves (Rand)	$RESZ_{t-1}$	1	10	15	16	0.0322	0.018

Table 7.4: Factor-Return Correlation Analysis (Continued…)

Panel F: Commodities							
Commodities (US$)	COM_t	10	15	1	11	0.264***	0.235***
Commodities (Rand)	$COMZ_t$	2	14	10	12	0.140*	0.121*
Non-fuel Commodities (US$)	NFC_t	10	16	0	10	0.217***	0.185**
Non-fuel Commodities (Rand)	$NFCZ_t$	1	10	15	16	0.061	0.003
Oil Prices (US$)	OIL_t	10	16	0	10	0.217***	0.228***
Oil Prices (Rand)	$OILZ_t$	2	21	3	5	0.175**	0.184**
Gold Prices (US$)	GLD_t	3	22	1	4	0.036	0.045
Gold Prices (Rand)	$GLDZ_t$	0	8	18	18	-0.0519	-0.037
Metal Prices	MET_t	16	10	0	16	0.208**	0.193***
Metal Prices (Rand)	$METZ_t$	2	19	5	7	0.123*	0.092
Panel G: Interest Rates							
Real Interest Rates	RIB_t	7	17	2	9	0.026	-0.007
3-Month T Bill Rates	$3TB_{t-1}$	0	21	5	5	-0.145**	-0.079
Long-Term Gov. Bond Yields	LTY_t	0	8	18	18	-0.044	-0.126*
Short-Term Gov. Bond Yields	STY_t	0	11	15	15	-0.089	-0.179**
Term Structure	TER_t	0	7	19	19	-0.113	-0.097
Panel H: Trade							
Trading Partner Lead. Index	TLI_t	25	1	0	25	0.497***	0.415***
Trading Partner Coinc. Index	TCI_{t-3}	13	13	0	13	0.122*	0.089
Terms of Trade	TOT_{t-2}	0	19	7	7	-0.192***	-0.253***
Panel I: Market Indices							
JSE All Share Index	R_{Mt}	26	0	0	26	-	-
MSCI World Index (US$)	R_{IMt}	26	0	0	26	0.684***	0.673***
MSCI World Index (Local)	R_{IMLt}	26	0	0	26	0.668***	0.663***
MSCI World Index (Rand)	R_{IMZt}	20	0	0	20	0.532***	0.481***
Panel J: Statistical Factors							
Factor 1	F_{1t}	22	4	0	22	0.283***	0.356***
Factor 2	F_{2t}	18	8	0	18	0.438***	0.415***
Factor 3	F_{3t}	20	6	0	20	0.636***	0.560***

Notes:
The asterisks, ***, ** and *, indicate statistical significance at the respective 1%, 5% and 10% levels of significance. Obs. refers to the number of observations. Notation refers to the formulaic notation used to abbreviate each factor. Pearson's (ordinary) correlation coefficient is denoted by ρ_P. Spearman's (rank-order) correlation coefficient is denoted by ρ_S. Sig. is the total number of significant ordinary correlations between the industrial sector return series and each factor series. ρ_P JSE is the ordinary correlation between returns on the JSE All Share Index and a given factor series. ρ_S JSE is the rank-order correlation between returns on the JSE All Share Index and a given factor series. The subscripts indicate the lag order for each factor that is significantly correlated with the greatest number of industrial sectors and the JSE All Share Index.

Although numerous factors show seemingly systematic correlation with the industrial sectors, for example, GFR_t, $NFCZ_t$ and TCI_{t-3}, only a handful qualify for incorporation into the linear factor model describing South African stock returns after the consideration of the relationship with returns on the JSE All Share Index and after controlling for structure breaks. These are (the innovations in) the number of building plans passed, BP_{t-1}, the domestic composite cyclical leading indicator $LEAD_{t-1}$, business activity, BUS_t, fluctuations in the Rand-Dollar exchange rate, USD_t, world metal prices, MET_t, long-term government bond yields, LTY_t and the leading indicator for South Africa's trading partners, TLI_t. This limited qualifying set demonstrates the difficulties in identifying relevant factors.

In the real activity category in Panel A of Table 7.4., unexpected changes in the number of building plans passed, BP_{t-1}, are positively correlated with 18 industrial sectors and returns on the JSE All Share Index. Given the widespread correlation with the industrial sectors and relatively high correlation with the JSE All Share Index, this factor is considered as a candidate factor for inclusion in the linear factor model specification. Szczygielski and Chipeta (2015; 13, 15) also find that this factor is significantly correlated with returns on the JSE All Share Index and that JSE All Share Index returns respond to innovations in this factor. The cyclical indicators that immediately garner attention in Panel C of Table 7.4. are the leading composite indicator, $LEAD_{t-1}$, and the business activity indicator, BUS_t. Unexpected changes in both indicators are positively and significantly correlated with over half of the return series for the industrial sectors although BUS_t appears to have a marginally more pervasive effect. This factor is correlated with 17 industrial sectors whereas $LEAD_{t-1}$ is positively correlated with 15 sectors. Both factors are significantly correlated with returns on the JSE All Share Index although BUS_t is significantly correlated with the market aggregate only when Spearman's rank correlation is considered (albeit weakly). An application of the Bai-Perron (1998) test does not indicate the presence of structural breaks.

In the exchange rate category in Panel D of Table 7.4., the correlation between the Rand-Dollar exchange rate, USD_t, and returns on the industrial sectors appears to be the most pervasive relative to the other factors. This factor is significantly and negatively correlated with 19 industrial sectors. With the exception of USD_t, which is negatively and significantly correlated (albeit weakly and correlation is limited to Spearman's correlation) with returns on the JSE All Share Index, none of the other exchange rate factors are significantly correlated with returns on the JSE All Share Index. This suggests that the other exchange rate factors are not systematic in nature. The Bai-Perron (1998) test for structural breaks is applied in breakpoint least squares single-factor models for all exchange rates in Panel D of Table 7.4. Results confirm that only USD_t has a pervasive impact, albeit of varying magnitude and direction, on returns on the South African stock market for the entire sample period. Consequently, USD_t is retained for further analysis. In the commodities category in Panel F of Table 7.4., the factor that is retained is MET_t, the unanticipated changes in the price of metals. This factor is positively and significantly correlated with returns on the JSE All Share Index and with 16 industrial sectors and can therefore be considered as pervasive.

In the interest rate category in Panel G of Table 7.4., the factor that is retained is the series of unanticipated changes in yields on long-term government bonds, LTY_t. This factor is negatively and significantly correlated with 18 industrial sectors and also correlated with returns on the JSE All Share Index, as evident from a statistically significant Spearman's correlation coefficient. This is the only factor in this category that appears to have a truly systematic impact after consideration is given to the presence of structural breaks in a breakpoint least squares specification. Trade factors in Panel H of Table 7.4. are assumed to represent economic conditions experienced by South Africa's trading partners. These conditions are (one of) the determinants of the foreign demand for South African goods and services (exports) (Baier & Bergstrand, 2001: 23; Vogt, 2008: 671).[115] The factor in this category that is widely correlated with both industrial sector returns and returns on the JSE All Share Index is the leading composite business cycle indicator for trading partner countries, TLI_t. This factor is positively correlated with almost all industrial sector return series and strongly correlated with returns on the JSE All Share Index. In preliminary analysis, the correlation coefficient between the changes in the terms of trade, TOT_{t-2}, and TLI_t, is statistically insignificant suggesting that TLI_t is more representative of global economic conditions rather than the direct impact of trade on the South African stock market. TLI_t is retained for further analysis.

Although all factors and categories are considered, for some categories, no factors are chosen. For other categories, certain factors appear to be potential candidate factors for inclusion in the linear factor model, but are disqualified following further consideration. In the prices category in Panel B of Table 7.4., no factors are considered as no factor is correlated with more than half of the industrial sector series. For the monetary factors in Panel E of Table 7.4., only one factor is considered. This is the level of foreign reserves denominated

[115] Both studies cited take into account the growth of global or trading partner income as a determinant of the demand for exports. This provides support for the hypothesis that global economic conditions are one of the factors that impact export demand. Both studies use GDP as a measure of income and therefore are a proxy for economic conditions. This study uses a measure of the terms of trade and coincident and leading composite business cycle indicators for trading partner countries. These measures are available on a monthly basis.

in South African Rand, $RESZ_{t-1}$. This factor is found not to be pervasive after accounting for structural breaks although it is highly correlated with one of the retained factors, USD_t (correlation of 0.801). For both reasons, it is excluded. In Panel F, the only factor that competes with the retained factor, MET_t, is the gold price denominated in Rands, $GLDZ_t$. This factor is excluded after further screening reveals that MET_t has a stronger correlation with the derived factor scores. Although in the interest rate category in Panel G of Table 7.4., changes in the short-term government bond yields, STY_t, and changes in the term structure, TER_t, are also correlated with over half of the industrial sector series, these factors are not correlated with returns on the JSE All Share index. This holds even after accounting for structural breaks. In the trade category in Panel H of Table 7.4., the coincident composite business cycle indicator for trading partner countries, TCI_{t-3}, is significantly correlated with half of the industrial sector return series and is weakly correlated with returns on the JSE All Share Index, as suggested by a significant ordinary correlation coefficient. However, an insignificant Spearman correlation coefficient suggests that TCI_{t-3} is uncorrelated with returns on the JSE All Share Index. As a result of this ambiguity, this factor is not retained.

Correlations between returns on the industrial sectors, returns on the JSE All Share Index and the three candidate MSCI World Market indices are reported in Panel I of Table 7.4. As expected, returns on the JSE All Share Index are significantly correlated with all industrial sectors. It remains to be determined which MSCI World Market Index will be used to derive the second residual market factor. The candidate factors are returns on the Dollar, local currency and Rand denominated MSCI World Market indices, denoted as R_{IMt}, R_{IMLt}, R_{IMZt}, respectively. Returns on all three indices are strongly correlated with returns on the JSE All Share Index. With the exception of R_{IMZt}, which is correlated with 20 industrial sectors, R_{IMt} and R_{IMLt} are correlated with all industrial sectors. All correlations are positive. The correlation between R_{IMLt} and returns on JSE All Share Index (0.668 (ordinary) and 0.663 (rank)) is marginally weaker relative to the correlation of R_{IMt} with returns on the JSE All Share Index (0.684 and 0.673). Of the three indices, R_{IMZt} shows the weakest correlation (0.532 and 0.481) with returns on the JSE All Share Index. The relatively strong correlation between R_{IMt} and returns on the JSE All Share Index, suggests that the informational content of the MSCI World Market Index denominated in US Dollars may have more to do with sentiment rather than changes in global economic fundamentals although global economic fundamentals will also be reflected in this index (Bradfield, 1990: 6).

On the basis of the strength of correlations, R_{IMt} , appears to be the most important formulation of the MSCI World Market Index for the South African stock market. To confirm that this is indeed the case, single-factor least squares specifications relating factor scores to R_{IMt}, R_{IMLt} and R_{IMZt} are estimated, following the form of equation (6.14). The $\bar{R}^2$ s and the associated significance of the coefficients are summarised in Table 7.5.

Table 7.5: Factor-MSCI World Market Index Return Regressions

Factors	R_{IMt}	R_{IMLt}	R_{IMZt}
F_{1t}	0.118***	0.116***	0.023**
F_{2t}	0.140***	0.118***	0.031**
F_{3t}	0.190***	0.202***	0.395***

Notes:
The asterisks, ***, ** and *, indicate statistical significance at the respective 1%, 5% and 10% levels of significance. Reported significance relates to the coefficient on the respective MSCI World Market Index in single-factor regressions. The reported values are the $\bar{R}^2$ for each factor regression. Least squares with Newey-West heteroscedasticity and autocorrelation consistent (HAC) standard errors used for estimation purposes.

From Table 7.5., it is immediately evident that R_{IMt}, R_{IMLt} and R_{IMZt} are important proxies for the pervasive influences reflected in the derived factor scores. The $\bar{R}^2$ for R_{IMt} is higher in the single-factor regressions of R_{IMt} on F_{1t} and F_{2t} relative to the $\bar{R}^2$ for R_{IMLt} and R_{IMZt}. This favours R_{IMt} over these two factors although the $\bar{R}^2$ is marginally higher for the regression of F_{3t} onto R_{IMLt} and much higher for the regression of F_{3t} onto R_{IMZt} which yields the highest $\bar{R}^2$ (0.395) for this factor. Nevertheless, as R_{IMLt} is correlated with 20 sectors (Panel J of Table 7.4.) and because of its poor ability to reflect the influences in F_{1t} and F_{2t}, as measured by the (relatively) low respective $\bar{R}^2$ s of 0.023 and 0.31, it is omitted for further analysis. This leaves R_{IMt} and R_{IMLt}. The correlation between these two factors is very high and highly significant; a correlation coefficient of 0.964 implies that these factors are similar and reflect similar influences. The use of either factor is likely to yield similar results. However, it appears that R_{IMt} is a marginally better proxy for F_{1t} and F_{2t} and only slightly underperforms R_{IMLt} for F_{3t} ($\bar{R}^2$ of 0.190 for R_{IMt} vs 0.202 for R_{IMLt}). Given that R_{IMt} appears to be a somewhat more suitable proxy for two of the three factors extracted from South African stock returns and is correlated with all industrial sectors and given that both factors are almost perfectly correlated, the second residual market factor, $IM\varepsilon_t$ is derived from this version of the index (the Dollar denominated MSCI World Market Index) (Section 6.4.2., equation (6.17)).

Finally, the derived statistical factors must also be examined to confirm that they represent systematic influences. This is especially pertinent given that these factors are derived from a limited sample of industrial sectors (Section 6.2.1.). This is done by examining the correlations between the statistical factors, F_{1t}, F_{2t} and F_{3t}, and returns on the JSE All Share Index, R_{Mt}. The correlations in Panel J of Table 7.4. show that all three factors are significantly correlated with R_{Mt}. F_{3t} and R_{Mt} exhibit the highest level of correlation with an ordinary correlation coefficient of 0.636. This declines for F_{2t} to 0.438 and to 0.283 for F_{1t}. Nevertheless, all correlations are highly statistically significant (*p*-values below 0.01) suggesting that these factors and movements in the JSE All Share index, which are assumed to broadly represent systematic influences, are related (Elton & Gruber, 1988: 40-42; Spyridis *et al.,* 2012: 52).

7.4. PROXIES FOR PERVASIVE INFLUENCES IN RETURNS

Having identified the candidate set of macroeconomic factors, namely BP_{t-1}, $LEAD_{t-1}$, BUS_t, USD_t, MET_t, LTY_t and TLI_t, and having specified the residual market factor, $M\varepsilon_t$, and the second residual market factor, $IM\varepsilon_t$, the next step is to confirm that these factors are proxies for the underlying influences in stock returns. The (original) correlation matrix for the macroeconomic factors and the two market factors (prior to orthogonalisation, denoted as R_{Mt} and R_{IMt} respectively) is presented in Table 7.6. As the residual market factors are uncorrelated with the remaining factors, they are omitted from Table 7.6. The correlation structure has the potential to distort the results of the estimated linear factor models if multicollinearity arises. The consequences of multicollinearity partly mimic those of underspecification; multicollinearity results in inflated standard errors, wider confidence intervals, unstable coefficients and coefficients of implausible magnitudes (Studenmund, 2014: 265-271; Williams *et al.,* 2013: 11).

Table 7.6: Correlation Matrix Of Retained Factors

	BP_{t-1}	$LEAD_{t-1}$	BUS_t	USD_t	MET_t	LTY_t	TLI_t	R_{Mt}	R_{IMt}
BP_{t-1}	1.000								
$LEAD_{t-1}$	0.155**	1.000							
BUS_t	0.079	-0.001	1.000						
USD_t	0.029	-0.027	-0.081	1.000					
MET_t	0.188***	0.039	0.074	-0.349***	1.000				
LTY_t	-0.031	0.126*	0.015	0.447***	-0.072	1.000			
TLI_t	0.193***	0.178**	0.138*	-0.215***	0.352***	0.016	1.000		
R_{Mt}	0.226***	0.279***	0.116	-0.114	0.208***	-0.044	0.497***	1.000	
R_{IMt}	0.127*	0.251***	0.101	-0.380***	0.214***	-0.120*	0.556***	0.684***	1.000

Notes:
The asterisks, ***, ** and *, indicate statistical significance at the respective 1%, 5% and 10% levels of significance. Correlation coefficients are ordinary (Pearson's) correlation coefficients. BP_{t-1} - (changes in) Building Plans Passed, $LEAD_{t-1}$ - (changes in) Leading Indicator, BUS_t - (changes in) Business Activity, USD_t - (changes in) Rand-Dollar Ex. Rate, MET_t - (changes in) Metal Prices, LTY_t - (changes in) Long-Term Gov. Bond Yields, TLI_t - Trading Partner Lead. Index, R_{Mt} - (returns on the) JSE All Share Index, R_{IMt} - (returns on the) MSCI World Index (US$).

Moreover, Mela and Kopalle (2002: 672) argue that depending upon the correlation structure, coefficient variance estimates may even decrease in the presence of increasing multicollinearity.[116] Therefore, it is desirable to assess the correlation matrix to determine whether possible multicollinearity may arise so that multicollinearity, the presence of underspecification and the application of remedial measures are not confounded.

The correlation coefficients in Table 7.6. point toward a number of statistically significant relationships. With the exception of the correlation between USD_t and LTY_t, TLI_t and R_{Mt}, TLI_t and R_{IMt}, and R_{Mt} and R_{IMt}, all correlation coefficients are below 0.4. However, there are a number of correlation coefficients above 0.2 that are statistically significant. While there is no defined level of correlation that is seen as problematic (aside from perfect correlation), Poon and Taylor (1991: 628) note that none of the correlation coefficients for the macroeconomic factors used in their application of the APT to the UK stock market are greater than 0.5. The authors suggest that correlations below this magnitude are not problematic. However, Bürki and Gaskell (2012: 620) are of the opinion that remedial measures should be taken for factors with correlations above 0.30 and Jaeger and Snider (2013: 63) state that remedial measures, in the form of orthogonalisation, should be taken whenever two factors are correlated.[117] Mela and Kopalle (2002: 667) identify various levels of correlation from literature as being seen as problematic; examples cited are 0.35, 0.7 and correlation as high as 0.9.

Following a number of preliminary regressions of the statistical factors onto the macroeconomic factor set following the functional form of equation (6.14), USD_t is orthogonalised against LTY_t and MET_t, and TLI_t is orthogonalised against MET_t. The resultant residual series, $USD\varepsilon_t$ and $TLI\varepsilon_t$, are used in place of the original innovation series. This selective orthogonalisation approach is followed to ensure that the structure of the correlation matrix in Table 7.6. is not significantly altered (for a discussion and criticism

---

[116] Mela and Kopalle (2002: 673) further show that the variance of estimated parameters is lower in a negatively correlated environment relative to an environment in which explanatory factors are uncorrelated or are positively correlated.

[117] It is worth noting that the papers of Bürki and Gaskell (2012) and Jaeger and Snider (2013) are in the discipline of cognitive sciences and may not be fully applicable to the economics discipline. Nevertheless, these papers (and the others cited in the text) demonstrate the divergence of opinions relating to the level of correlation that calls for remedial action.

of this approach, see Wurm & Fisicaro, 2014: 41).[118] By construction, $USD\varepsilon_t$ is no longer correlated with LTY_t and MET_t , and $TLI\varepsilon_t$ is no longer correlated with MET_t. It also follows that the two residual market factors, $M\varepsilon_t$ and $IM\varepsilon_t$, are uncorrelated with each other and all remaining macroeconomic factors (equation (6.15); equation (6.17); Section 6.3.2.).

Following the preceding analysis of the factor correlation structure, factor regressions are performed following the functional form of equation (6.14) (incorporating innovations in the macroeconomic factors), equation (6.16) (incorporating innovations in macroeconomic factors and the domestic residual market factor, $M\varepsilon_t$), and finally, equation (6.18) (incorporating innovations in macroeconomic factors and both residual market factors, $M\varepsilon_t$ and $IM\varepsilon_t$):

$$F_{nt} = \alpha + b_{nBP}BP_{t-1} + b_{nLEAD}LEAD_{t-1} + b_{nBUS}BUS_t + b_{nUSD\varepsilon}USD\varepsilon_t + b_{nMET}MET_t + b_{nLTY}LTY_t + b_{nTLI}TLI\varepsilon_t + \varepsilon_{nt} \tag{7.1}$$

$$F_{nt} = \alpha + b_{nBP}BP_{t-1} + b_{nLEAD}LEAD_{t-1} + b_{nBUS}BUS_t + b_{nUSD\varepsilon}USD\varepsilon_t + b_{nMET}MET_t + b_{nLTY}LTY_t + b_{nTLI}TLI\varepsilon_t + b_{nM\varepsilon}M\varepsilon_t + \varepsilon_{nt} \tag{7.2}$$

$$F_{nt} = \alpha + b_{nBP}BP_{t-1} + b_{nLEAD}LEAD_{t-1} + b_{nBUS}BUS_t + b_{nUSD\varepsilon}USD\varepsilon_t + b_{nMET}MET_t + b_{nLTY}LTY_t + b_{nTLI}TLI\varepsilon_t + b_{nM\varepsilon}M\varepsilon_t + b_{IM\varepsilon}IM\varepsilon_t + \varepsilon_{nt} \tag{7.3}$$

where in equation (7.1), (7.2) and (7.3), F_{nt} is statistical factor *k*, the respective betas, b_{nk} s, represent the sensitivity of the factor scores of factor *k* to the macroeconomic factor set,

BP_{t-1}, $LEAD_{t-1}$, BUS_t, $USD\varepsilon_t$, MET_t, LTY_t and $TLI\varepsilon_t$ in equation (7.1) and to the residual market factor derived from returns on the JSE All Share Index, $M\varepsilon_t$ in equation (7.2) and also the second residual market factor derived from returns on the US Dollar denominated MSCI World Market Index, $IM\varepsilon_t$, in equation (7.3). The estimation of these specifications serves a dual purpose. The first aligns with the primary purpose of this chapter, which is to explore the factor structure of South African stock returns and to confirm whether the factors that have been identified are indeed relevant to the South African stock market and can describe the return generating process by proxying for the pervasive influences in stock

[118] The correlation between the original innovation series and the orthogonalised series for USD_t and $USD\varepsilon_t$ is 0.836 and for TLI_t and $TLI\varepsilon_t$, the correlation is 0.936 suggesting that orthogonalisation does not substantially change the nature of these factors.

returns. The second, which aligns more with the purpose of Chapter 8, Chapter 9 and Chapter 10, is to determine how well the macroeconomic factor set and $M\varepsilon_t$ and $IM\varepsilon_t$ approximate the pervasive factors underlying the South African stock market. A poor ability of the factor set, even when $M\varepsilon_t$ and $IM\varepsilon_t$ are included as proxies for omitted domestic and global factors, to proxy for F_{1t}, F_{2t} and F_{3t} is a preliminary indication that a linear factor model characterised by macroeconomic factors and the two residual market factors is unable to adequately describe the return generating process.

Although the inclusion of seven macroeconomic factors in equation (7.1) yields an unparsimonious specification, Middleton and Satchell (2001: 506) advocate for specifications to include a "generous" number of pre-specified proxy macroeconomic factors that is equal to or greater than the number of true factors. By following this approach, it is hoped that the likelihood of underspecification associated with the use of macroeconomic factors, as noted by Van Rensburg (2000: 36) and Middleton and Satchell (2001: 506), is reduced. It then remains for it to be determined whether the residual market factors can resolve any remaining underspecification.

Table 7.7. presents the results of the factor score regressions. All macroeconomic factors and the residual market factors are statistically significant across the factor regressions in Panel A, Panel B and Panel C. In the factor regression for F_{1t} across the panels, all factors with the exception of MET_t and $TLI\varepsilon_t$ are statistically significant. For F_{2t} , MET_t and $TLI\varepsilon_t$ are statistically significant across the panels. For F_{3t}, $LEAD_{t-1}$, $USD\varepsilon_t$, LTY_t and $TLI\varepsilon_t$ are statistically significant in Panel B and Panel C. In Panel B (equation 7.2), the residual market factor is statistically significant for the factor regressions of all factor scores onto the factor set. In Panel C (equation 7.3), the international residual market factor is statistically significant in the factor regressions of F_{1t} and F_{3t}. This confirms that these factors, namely the macroeconomic factors and the residual market factors, are proxies for the pervasive influences in returns and therefore, all factors are retained. The *F*-statistic is statistically significant across specifications, confirming the overall significance of the approximation of the factor scores consisting of the macroeconomic factors and both residual market factors.

Table 7.7: Factor Score Regressions

	Panel A: Restricted Model			Panel B: Unrestricted Market Model			Panel C: Unrestricted Model		
Factor	F_{1t}	F_{2t}	F_{3t}	F_{1t}	F_{2t}	F_{3t}	F_{1t}	F_{2t}	F_{3t}
Intercept	0.007	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
BP_{t-1}	1.339**	-0.364	0.401	1.339**	-0.364	0.401	1.339**	-0.364	0.401
$LEAD_{t-1}$	15.467*	7.756	15.460*	15.467*	7.756	15.460**	15.467*	7.756	15.460**
BUS_t	1.661**	0.766	0.565	1.661**	0.766	0.565	1.661**	0.766	0.565
$USD\varepsilon_t$	-7.763***	-3.621	6.531***	-7.763***	-3.621	6.531***	-7.763***	-3.621	6.531***
MET_t	-1.479	8.713***	1.372	-1.479	8.713***	1.372	-1.479	8.713***	1.372
LTY_t	-146.237***	-28.506	40.792	-146.237***	-28.506	40.792*	-146.237***	-28.506	40.792*
$TLI\varepsilon_t$	10.485	48.334*	82.364***	10.485	48.334**	82.364***	10.485	48.334**	82.364***
$M\varepsilon_t$				5.430***	9.081***	15.897***	5.430***	9.081***	15.897***
$IM\varepsilon_t$							4.317*	1.174	3.918*
$\bar{R}^2$	0.246	0.161	0.131	0.285	0.259	0.431	0.294	0.255	0.437
AIC	2.727	2.997	3.057	2.679	2.879	2.637	2.671	2.888	2.632
F-statistic	11.962***	2.641**	6.555***	11.690***	4.695***	21.852***	12.739***	4.570***	18.458***

The asterisks, ***, ** and *, indicate statistical significance at the respective 1%, 5% and 10% levels of significance. Least squares with Newey-West heteroscedasticity and autocorrelation consistent (HAC) standard errors used for estimation purposes. All factors are in innovations (unexpected changes) (Section 6.2.2; Section 6.2.3.; Table 6.4.), where BP_{t-1} - Building Plans Passed, $LEAD_{t-1}$ - Leading Indicator, BUS_t - Business Activity, USD_t - Rand-Dollar Ex. Rate, MET_t - Metal Prices, LTY_t - Long-Term Gov. Bond Yields, TLI_t - Trading Partner Lead. Index, R_{Mt} - JSE All Share Index and R_{IMt} - MSCI World Index (US$).

Perhaps expectedly, the results indicate that the unrestricted factor regressions in Panel C in Table 7.7. provide the best approximation of the statistical factors. The $\bar{R}^2$ for F_{1t} is 0.294, for F_{2t} it is 0.255 and for F_{3t} it is 0.437. The relatively high $\bar{R}^2$s, with the exception of F_{2t}, which has a marginally higher $\bar{R}^2$ in Panel B relative to that in Panel C, suggest that the unrestricted model provides, for F_{1t} and F_{3t}, the best although incomplete approximation of these factors. That the $\bar{R}^2$ values for each factor regression are far below one attests to this. That the unrestricted model is the most appropriate representation of F_{1t} and F_{3t} is confirmed by the AIC values, which are lower relative to those of the specifications reported in Panel A and Panel B. The exception and as with the $\bar{R}^2$, is F_{2t} in Panel B, which has a (marginally) lower AIC value relative to that for the corresponding factor regression in Panel C. This suggests that for this factor, the inclusion of $IM\varepsilon_t$ in Panel C in the factor set does not improve the ability of the factor set to approximate the pervasive influences reflected by this factor. In summary, the unrestricted model appears to provide the best proxy for F_{1t} and F_{3t} relative to the restricted specifications in Panel A and Panel B of Table 7.7., and a relatively adequate proxy for F_{2t}. The inability of the macroeconomic factor set and $M\varepsilon_t$ and $IM\varepsilon_t$ to fully approximate the statistical factors, as suggested by an $\bar{R}^2$ that is far below one, indicates that these factors may be unable to fully explain the return generating process and therefore, the linear factor model may be underspecified.

The residual market factors improve the approximation of the statistical factors although the contribution of $IM\varepsilon_t$ is marginal. The $\bar{R}^2$ for F_{1t} in Panel A (the restricted model comprising macroeconomic factors) is 0.246, 0.161 for F_{2t} and 0.131 for F_{3t} respectively. This indicates that on their own, the macroeconomic factors are relatively poor proxies for the pervasive factors in returns. The inclusion of the domestic residual market, $M\varepsilon_t$, in Panel B substantially improves the approximation of the statistical factors. The $\bar{R}^2$ for F_{1t} increases to 0.285 from 0.246 and the increase in the $\bar{R}^2$s is most notable for F_{2t} and F_{3t}. For these latter factors, the $\bar{R}^2$ increases to 0.259 from 0.161 for F_{2t} and to 0.431 from 0.131 for F_{3t}. For all three factor regressions in Panel B, the corresponding AIC values decrease, thereby confirming the improved model fit. While the inclusion of $M\varepsilon_t$ improves the approximation of the underlying factors, as represented by the factor scores, the improvement is not significant enough to closely approximate these factors. This calls into question the early assertion made by Berry *et al.* (1988: 31), namely that the "worry over possible missing factors is substantially resolved by using a residual market factor" (Section 3.4). It appears that this concern still remains, given the low $\bar{R}^2$ values. The finding that $M\varepsilon_t$ is statistically significant across the factor regressions in Panel B and that $IM\varepsilon_t$ is statistically significant for F_{1t} and F_{3t} in Panel C suggests that the macroeconomic factors by themselves do not account for the pervasive influences in stock returns. Therefore, it appears that having an equal or greater number of proxy factors for the true number of factors does not guarantee a comprehensive approximation of the underlying factors in returns (Middleton & Satchell, 2001: 506). In the present case, the hypothesised true number of factors , as determined by factor analysis, is three. However, as suggested by the relatively low $\bar{R}^2$s, it appears that the eight and nine factor structures in Panel B and Panel C of Table 7.7. are unable to adequately approximate these three factors. Additionally, the analysis of the results in Panel A supports Van Rensburg's (2000: 36) contention that specifications that employ pre-specified macroeconomic factors may suffer from underspecification.

The inclusion of $IM\varepsilon_t$ translates into a marginal improvement in the $\bar{R}^2$ for F_{1t} and F_{3t} in the unrestricted model and this factor is statistically significant in the respective factor regressions. The $\bar{R}^2$ for F_{1t} increases marginally from 0.285 in Panel B to 0.294 in Panel C and from 0.431 for F_{3t} in Panel B to 0.437 in Panel C. This suggests that most of the information that is reflected in the (unorthogonalised) returns on the international market index, R_{IMt}, is already reflected in the macroeconomic factors and $M\varepsilon_t$. For this reason, the contribution of $IM\varepsilon_t$ is marginal. This can be attributed to the type of factors included in the factor set. For example, the leading trading partner indicator index, $TLI\varepsilon_t$, is indicative of the economic conditions experienced by South Africa's trading partners (excluding the influence of world metal prices). As such, this factor may account for some of the global influences in stock returns that would otherwise have been reflected in R_{IMt} in the absence of this factor and before orthogonalisation. However, a finding of a marginal contribution does not invalidate the potential role of a second residual market factor. $IM\varepsilon_t$ is statistically significant in the factor regressions for F_{1t} and F_{3t} in Panel C. This suggests that it reflects information that should be reflected in the conventional residual market factor, if the residual market factor is indeed a proxy for all omitted factors (Chang, 1991: 380; Kryzanowski *et al.,* 1994: 155-156). The significance of $IM\varepsilon_t$ in Panel C for the two abovementioned factors challenges the assertion that the conventional residual market factor is an adequate proxy for omitted factors. This is further investigated in the empirical analysis in the chapters that follow.

7.5. CHAPTER SUMMARY AND CONCLUSION

This chapter investigates the factor structure underlying the South African stock market. The factor structure is characterised by strong interdependence between the return series. Three statistical factors are extracted from returns (Section 7.2.). To identify macroeconomic factors related to returns on the industrial sectors that comprise the sample and are proxies for the pervasive influence in stock returns, factor-return correlation analysis is undertaken in Section 7.3. This involves a preliminary screening of an extensive number of domestic and globally orientated factors categorised as representative of real activity, prices, cyclical indicators, exchange rates, monetary factors, commodity prices, interest rates and trade. Factors that show widespread correlation with the return series and are significantly correlated with returns on the JSE All Share Index are deemed to have a pervasive impact and are retained for further analysis. That only seven macroeconomic factors qualify, namely BP_{t-1}, $LEAD_{t-1}$, BUS_t, USD_t, MET_t, LTY_t and TLI_t, is indicative of the challenges associated with identifying macroeconomic factors that proxy for pervasive influences in returns. Following correlation analysis, USD_t and TLI_t are orthogonalised and the orthogonal versions of these factors are used, $USD\varepsilon_t$ and $TLI\varepsilon_t$, in subsequent factor regressions.

In Section 7.4., factor regressions show that the macroeconomic factors retained for further analysis are related to the factor scores derived from the return series. However, macroeconomic factors appear to be poor proxies for the underlying influences in returns. This is suggested by the low resultant $\bar{R}^2$ s . It also appears that $M\varepsilon_t$ is an incomplete proxy for underlying pervasive factors. This is suggested by an $\bar{R}^2$ that is below one for the factor set incorporated into the unrestricted market model and the significance (as opposed to the redundancy) of $IM\varepsilon_t$ in the unrestricted specification.

The poor ability of the macroeconomic factors to proxy for influences in South African stock returns points towards possible underspecification that may not be resolved by the use of the residual market factors. This is the subject of the chapters that follow. Chapter 8 constructs a benchmark model with the intention of specifying a model that is generally free of the underspecification caused by the omission of pervasive factors. Chapter 9 investigates the potential consequences of underspecification in a linear factor model that comprises the identified macroeconomic factors. Chapter 10 considers the efficacy of the residual market factor in resolving factor omission and the associated consequences of underspecification.

CHAPTER 8

THE BENCHMARK MODEL

8.1. INTRODUCTION

This chapter develops the benchmark model described in Section 6.4.1. upon the basis of the factor structure explored in Chapter 7. As this model relates industrial sector returns to the macroeconomic factors identified in the preceding chapter, the results are interpretable and provide insight into the macroeconomic forces impacting the South African stock market. It is also the specification against which the restricted and unrestricted models set out in Section 6.4.1. are compared. The focus of this chapter is a theoretically optimal and econometrically accurate description of the return generating process.
The residual correlation (analogously covariance) matrix is also analysed, to confirm that the linear model is correctly specified in that no omitted factors are reflected in the residual correlation matrix. Consequently, this chapter considers the validity of the assumption of uncorrelated residuals across return series (equation (2.2); Section 2.2.). Importantly, this chapter seeks to present a model that is free of underspecification attributable to the omission of pervasive factors. The benchmark specification provides a benchmark model and a residual correlation structure against which the ability of macroeconomic factors, and later on, the ability of the residual market factors to adequately reflect pervasive influences, can be assessed. It does so by incorporating a factor analytic augmentation that is deemed to reflect all influences that are not reflected in the macroeconomic factor set and the two residual market factors. The existence of such a factor analytic augmentation in the first place indicates that there may be other influences that are not reflected in the factor set.

This chapter proceeds as follows; Section 8.2. outlines the model specification and formulation and Section 8.3. summarises the results and interrogates the economic rationale of the model. Section 8.4. reports the model diagnostics and Section 8.5. investigates residual variance and the underlying conditional variance structures. The predictive ability of the model is considered in Section 8.6., and Section 8.7. directly investigates and tests for factor omission. Section 8.8. investigates the structure of the resultant residual correlation matrix. The primary aim of the tests for factor omission and the analysis of the residual correlation matrix is to establish the validity of the benchmark model. Section 8.9. concludes the chapter and summarises key findings.

8.2. BENCHMARK MODEL SPECIFICATION

To obtain the residuals for each industrial sector series and to derive the residual correlation matrix for the factor analytic augmentation in the benchmark model, the unrestricted model is first estimated. Equation (6.19) is now respecified in terms of the macroeconomic factors that are identified as proxies for pervasive factors in returns in Section 7.4. and the two residual market factors. This yields the following specification:

$$\begin{aligned} R_{it} &= \alpha + b_{iBP}BP_{t-1} + b_{iLEAD}LEAD_{t-1} + b_{iBUS}BUS_t + b_{iUSD\varepsilon}USD\varepsilon_t + b_{iMET}MET_t + b_{iLTY}LTY_t \\ &\quad + b_{iTLI}TLI\varepsilon_t + b_{iM\varepsilon}M\varepsilon_t + b_{iIM\varepsilon}IM\varepsilon_t + \varepsilon_{it} \end{aligned} \tag{8.1}$$

where R_{it} is the return on industrial sector index *i* at time *t*, the b_i's are the sensitivities to innovations in the respective macroeconomic factors, namely (as before in Section 7.4.) the number of building plans passed (BP_{t-1}), the leading composite (domestic) business cycle indicator ($LEAD_{t-1}$), domestic business activity (BUS_t), the (orthogonalised) Rand Dollar exchange rate ($USD\varepsilon_t$), metal prices in US Dollars (MET_t), long-term interest rates (LTY_t), the (orthogonalised) composite leading economic conditions indicator for South Africa's trading partners ($TLI\varepsilon_t$), the residual market factor derived from returns on the JSE All Share Index ($M\varepsilon_t$) and the second residual market factor derived from returns on the MSCI World Market Index ($IM\varepsilon_t$). Least squares estimation is applied in equation (8.1) to generate the residuals for factor analysis and the interpretation of the results the unrestricted model are not of interest at this stage.

The residuals of equation (8.1) are factor analysed and this produces two factor analytically derived factors. As in Section 7.2. , Velicer's (1976) MAP test is applied to determine the number of factors to extract. The eigenvalues for the first and second factor indicate that the first factor explains 21.27% of the proportion of variance in the residuals of the unrestricted model whereas the second factor accounts for 9.13%. The scree test also points towards two common factors in the residuals of the unrestricted model and indicates that factors beyond the second factor are pseudofactors. Such factors are those that can explain some of the asset -specific variance for a few sectors, factors that are trivial or factors that are non-trivial but not general in that they are important for specific subsets or during specific periods of time (Kryzanowksi & To, 1983: 39; Connor, 1995: 44).[119] Consequently, two factors are extracted and appended to the unrestricted specification to obtain the benchmark model, denoted by equation (8.2):

$$\begin{aligned} R_{it} &= \alpha + b_{iBP}BP_{t-1} + b_{iLEAD}LEAD_{t-1} + b_{iBUS}BUS_t + b_{iUSD\varepsilon}USD\varepsilon_t + b_{iMET}MET_t + b_{iLTY}LTY_t \\ &+ b_{iTLI}TLI\varepsilon_t + b_{iM\varepsilon}M\varepsilon_t + b_{iIM\varepsilon}IM\varepsilon_t + b_{i1}f_{1t} + b_{i2}f_{2t} + \varepsilon_{it}^{*} \end{aligned} \tag{8.2}$$

where all factors are as before with the exception of the two statistical factors, f_1 and f_2, which are factor scores derived from the residuals of equation (8.1), and their associated coefficients, b_{i1} and b_{i2}, respectively. These two factors comprise the factor analytic augmentation. The theoretically purely idiosyncratic component is represented by ε_{it}^{*}. As inference and estimate efficiency are of interest, equation (8.2) is estimated using ML estimation with ARCH/GARCH errors as outlined in Section 6.4.2.

8.3. RESULTS AND MODEL OVERVIEW

8.3.1. Macroeconomic Factor Significance And Economic Interpretation

The abridged results of the benchmark model are presented in Table 8.1. The results in Panel A indicate that 119 of the 182 (65.38%) estimated coefficients for the seven macroeconomic factors are statistically significant. All coefficients on the conventional residual market factor are statistically significant and the total number of significant coefficient estimates increases to 145 out of 208 (69.71%) when the macroeconomic factors and the residual market factor are considered together. When considered with the second residual market factor derived from the MSCI World Market Index, 161 of the 234 (68.80%) estimated coefficients are statistically significant.

The number of observed significant coefficients and the corresponding percentages permit a comparison to other studies that also apply a single linear factor model specification across an extended number of series. In McElroy and Burmeister (1988: 36-37), 171 of the 280 (61.11%) estimated coefficients are statistically significant for a model relating returns on 70 individual stocks from the CRSP database to the four macroeconomic factors in the specification, namely the term structure of interest rates, the default spread, unexpected

[119] To test whether these factors are common over time, factor analysis is conducted on two equal subperiods, 2001M01 to 2008M12 and 2009M01 to 2016M12. For both subperiods, two factors are extracted upon the basis of the MAP test. This provides further evidence that the two factors incorporated into the benchmark model are not pseudofactors as they continue to persist across subperiods.

inflation and real final sales. When considered with the residual market factor derived from the S&P 500, 241 of the 350 (68.86%) estimated coefficients are statistically significant.[120]

Table 8.1: Summary Of Benchmark Model Results

Panel A: Coefficient And Significance Summary

Factor	Mean Coeff. Std Error |Z-score|	Mean LS Co. |Diff.|	$b_{ik} > 0$	$b_{ik} = 0$	$b_{ik} < 0$	Total Sig.
Intercept	0.006 (0.003) [2.013]	0.006 0.0006	13	13	-	13
BP_{t-1}	0.037 (0.031) [1.474]	0.038 0.001	10	16	-	10
$LEAD_{t-1}$	0.905 (0.425) [2.449]	0.883 0.022	17	9	-	17
BUS_t	0.079 (0.038) [2.190]	0.079 0.0006	18	8	-	18
$USD\varepsilon_t$	-0.180 (0.114) [2.075]	-0.174 0.006^{W}	1	10	15	16
MET_t	0.155 (0.074) [2.410]	0.173 0.018^{*W}	13	12	1	14
LTY_t	-3.920 (1.124) [4.190]	-3.727 0.193	-	6	20	20
$TLI\varepsilon_t$	2.865 (0.824) [3.809]	3.076 0.211*	24	2	-	24
$M\varepsilon_t$	0.664 (0.087) [9.078]	0.679 0.015	26	-	-	26
$IM\varepsilon_t$	0.217 (0.122) [2.435]	0.222 0.005	15	-	1	16
f_{1t}	-		16	-	4	20
f_{2t}	-		20			20

Panel B: Goodness-of-fit And Model Selection Criteria

	Mean	Minimum	Maximum
$\bar{R}^2$	0.504	0.171 Fixed line telecom.	0.941 Mining
AIC	-3.348	-4.956 Mining	-2.035 Fixed line telecom.
BIC	-3.114	-4.736 Mining	-1.797 Fixed line telecom.

[120] A two-tailed test and a 10% level of significance assumed when summarising the number of statistically significant coefficients.

Table 8.1: Summary Of Benchmark Model Results (Continued…)

Notes:
The asterisks, ***, ** and *, indicate statistical significance at the respective 1%, 5% and 10% levels of significance. All factors are in innovations (unexpected changes) (Section 6.2.2; Section 6.2.3.; Table 6.4.), where BP_{t-1} - Building Plans Passed, $LEAD_{t-1}$ - Leading Indicator, BUS_t - Business Activity, USD_t - Rand-Dollar Ex. Rate, MET_t - Metal Prices, LTY_t - Long-Term Gov. Bond Yields, TLI_t - Trading Partner Lead. Index, R_{Mt} - JSE All Share Index and R_{IMt} - MSCI World Index (US$). In Panel A, Mean Coeff. is the mean value of the intercept and the coefficients associated with each factor. Values in the parentheses ( ) are the mean coefficient standard errors (Std Error) and the values in the brackets [ ] are the mean *z*-scores (|Z-score|). In the third column, Mean LS Co. are the mean values of least squares intercepts and coefficients of the benchmark model. |Diff.| are the absolute mean differences between ML and least squares coefficients. $b_{ik} > 0$ and $b_{ik} < 0$ indicate the respective number of coefficients that are statistically significant and have a positive or negative impact. Total Sig. is the total number of statistically significant coefficients associated with each factor across the return series in the sample. In Panel B, Mean is the arithmetic mean of the $\bar{R}^2$, AIC and BIC values across sectors. The Minimum and Maximum values correspond to the lowest and highest values observed and the associated sectors for which they are observed. Throughout, the superscript "W" indicates that the Wilcoxon matched-pairs signed-rank test contradicts the results of the paired-sample *t*-test.

Berry *et al.* (1988: 3) find that 233 of the 316 (73.74%) estimated coefficients are statistically significant for a linear factor model relating returns on a sample of 79 US industrial sectors to macroeconomic factors in the linear factor model, namely default risk, the term structure, inflation/deflation and unexpected changes in the growth of profits (as measured by real final sales). When considered with the residual market factor derived from returns on the S&P 500, 312 of the 395 (78.99%) estimated coefficients are statistically significant. Although the present linear factor model specification is more complex relative to the five-factor specifications in these studies and is estimated for a different market, the percentages of significant coefficient estimates are closely comparable to those in McElroy and Burmeister (1988) and somewhat comparable to those in Berry *et al.* (1988). The preceding comparison indicates that the current linear factor model has descriptive validity for the South African stock market. However, it is possible that the descriptive validity is somewhat weaker in the absence of the factor analytic augmentation employed to estimate the benchmark model but not used by the studies cited above (Bilson *et al.,* 2001).[121]

[121] The findings of Bilson *et al.* (2001: 412-413) suggest that APT-type multifactor models may not have descriptive validity for all markets. For example, the authors find that the $\bar{R}^2$ for a five-factor model incorporating returns on the MSCI World Index, the money supply, consumer prices, industrial production and exchange rates is either 0 or close to zero for seven developing markets in the sample, namely Argentina, Brazil, Colombia, Greece, India, Jordan and Venezuela. The null hypothesis that all coefficients are jointly equal to zero is not rejected for these markets. Although it is possible that the specific factors used by the authors may not sufficiently describe the return generating process for certain markets (this is addressed by the authors), it demonstrates that a given specification may not have descriptive validity. The results of the present study are compared to those of McElroy and Burmeister (1988) and Berry *et al.* (1988) as these are seminal studies, retain close theoretical proximity to the APT framework and are also granular (conducted at individual firm level and industrial sector level respectively).

The analysis of the significance and direction of impact of the macroeconomic factors that are assumed to proxy for the pervasive influences in stock returns permits an interpretation of the results and a confirmation that the estimated relationships meet *a priori* expectations.

Innovations in the number of building plans passed, BP_{t-1}, have an overall positive impact on returns with a mean coefficient of 0.037. Moolman (2003: 295) considers this factor to be a macroeconomic indicator that reflects the expectations of economic agents. A positive impact on returns is therefore expected; an unanticipated improvement (deterioration) in expectations relating to the macroeconomic environment will translate into increased (decreased) expected future cash flows. Szczygielski and Chipeta (2015: 15) also find that this factor has a positive and statistically significant impact on returns on the JSE All Share Index. This contrasts with Van Rensburg's (1996: 106) preliminary findings, which suggest that this factor does not have an impact on South African stock returns. The relatively limited significance of this factor in the benchmark specification contrasts with the widespread significant correlation between BP_{t-1} and 18 industrial sectors reported in Table 7.4. The finding that this factor has a limited impact can be considered together with the correlation matrix reported in Table 7.6., which indicates that BP_{t-1} is significantly correlated with $LEAD_{t-1}$, MET_t and also TLI_t suggesting that the lack of widespread significance in the linear factor model may be attributable to multicollinearity. It is possible that although this factor has a widespread significant impact on the returns in the sample, its impact is weak and is subsumed in a multifactor setting.

Moolman (2003: 296-297) investigates the ability of a set of macroeconomic factors and indicators to predict turning points in the business cycle. Amongst these is the number of building plans passed as well as the composite leading indicator, $LEAD_{t-1}$. Using a probit model and with reference to the pseudo $\bar{R}^2$, it is shown that that BP_{t-1} has some predictive power in forecasting turning points in the business cycle for between 10 and 18 months ahead.[122] This is in contrast to the leading indicator, $LEAD_{t-1}$, which is found to have the greatest predictive power three months ahead, suggesting that it reflects short-term expectations relating to turning points in the business cycle and the associated changing

[122] Moolman (2003: 297) reports that the highest pseudo $\bar{R}^2$ is for 16 months and ranges between 0.227 (at 10 months ahead) and 0.302 (at 16 months ahead).

economic conditions (Venter, 2005).[123] In contrast to BP_{t-1}, this factor's favourable performance in the benchmark model is evident from a significant and positive relationship with 17 industrial sectors. The mean coefficient for $LEAD_{t-1}$ is 0.905. An advantageous characteristic of this factor is that while it comprises a broad and comprehensive range of components, it is readily interpretable, meaningful and easy to monitor (see Venter & Pretorius, 2004: 68 for a discussion of composition). Niemira (1993: 364) argues that the impact of leading indicators on stock prices is through an anticipation of changes in earnings. Anticipated changes in the business cycle, such as an end of a recession or the nearing of a recession, lead to changes in the sales cycle suggesting an associated improvement or deterioration in future earnings. The discrepancy between the ability of BP_{t-1} and $LEAD_{t-1}$ to explain returns is potentially related to investors placing a greater focus on short-term expectations. Investors appear to be more concerned with short-term expectations and near-term changes in the economic state, reflected by $LEAD_{t-1}$, relative to long-term expectations, reflected by BP_{t-1} (Pilinkus, 2010).[124]

Unanticipated changes in business activity, BUS_t, are derived from a constituent of the Purchasing Managers' Index compiled by the Bureau for Economic Research (BER) (2015).[125] The mean coefficient on this factor is 0.079 and there is a positive relationship between BUS_t and returns on 18 industrial sectors. The BER (2015: 5) measures business activity by production volumes, units of work accomplished, person-hours worked, sales volumes and other non-monetary measures. Similarly to BP_{t-1} and $LEAD_{t-1}$, this factor may be viewed as a composite index with a specific interpretation. However, the literature indicates that its function is different. Harris (1991: 65, 67), using US manufacturing output, reports that the PMI is not a reliable leading indicator of turning points nor does it signal changing

[123] Moolman (2003: 297) reports that the highest pseudo $\bar{R}^2$ is at three months and ranges between 0.713 (at three months ahead) and declines gradually to 0.310 at 10 months ahead.

[124] It is worth noting that the number of building plans passed is a component of the composite leading business cycle indicator. A weak but statistically significant correlation between these two factors (of 0.155) is observed in Table 7.6. Moolman (2003: 293) reports that the composite leading business cycle indicator is one of the better predictors of turning points in the South African business cycle.

[125] The Purchasing Managers Index (PMI) is compiled by the Bureau for Economic Research (BER) and is sponsored by ABSA, a large South African bank. The BER is located at the University of Stellenbosch. This version of the PMI is based upon the PMI produced by the Institute for Supply Management (ISI) in the US. The PMI is compiled on a monthly basis with a focus on the manufacturing sector and is considered a general indicator of business conditions (BER, 2015: 1).

economic trends. Instead, the PMI and its constituents represent an "imperfect but useful addition to our knowledge of current (emphasis added) economic conditions."[126] Kauffman (1999: 34-35) examines the relationship between constituents of the PMI and US GNP. The production (analogous to business activity) and new orders components of the PMI coincide with and are highly correlated (0.900 and 0.888) with GNP growth. The complementary usefulness of a business activity measure (to that of BP_{t-1} and $LEAD_{t-1}$) is apparent; BUS_t captures coincident changes in the macroeconomic state, which are reflected in less frequent measures of economic activity, namely the GDP and GNP. Kauffman (1999) further states that this data may be used to assess the effects of current economic conditions and that production is a good indicator of general economic conditions. Wongbangpo and Sharma (2002: 30) argue that the level of real economic activity, as measured by the GNP and proxied by BUS_t in the present model, is likely to influence stock prices through an impact on corporate profitability. It is postulated that an increase (decrease) in output will increase (decrease) expected future cash flows and thereby raise (lower) stock prices. The positive relationship observed between returns and BUS_t provides support for this proposed transmission mechanism.

The impact of exchange rates on stock prices is widely studied. For example, Griffin and Stultz (2001: 223, 225) investigate the impact of the foreign exchange rate, as measured by the Yen-Dollar exchange rate, on 17 Japanese traded goods industrial sectors. Most industries are found to be negatively impacted by a depreciation in the exchange rate. The two industries that are positively impacted are the integrated oil and steel industrial sectors. The authors state that this is consistent with exporting industries losing from an appreciation in the currency and importing industries benefitting from a currency depreciation. Pan, Fok and Liu (2007: 503-504) outline a number of transmission mechanisms between exchange rate fluctuations and stock prices. As in Griffin and Stultz (2001), it is suggested that a depreciation in the currency improves the competitiveness of exporting firms and increases foreign demand and sales, benefitting exporting firms and export orientated industries. For importing firms, an appreciation (depreciation) in the domestic currency translates into an increase (decrease) in the firm's receivables or accounts payable denominated in a foreign

[126] The version of the PMI discussed by Harris (1991) has roughly the same components as the BER PMI. For the version discussed by Harris (1991), manufacturing production may be seen as analogous to business activity in BER's PMI. BER (2015:1) suggests that the business activity component is a measure of business output.

currency and thereby increases (decreases) future profits. Also, firms may be impacted by changes in input prices driven by fluctuations in exchange rates. A depreciation (appreciation) of the exchange rate will increase (decrease) imported input costs and decrease (increase) expected future cash flows. Finally, it is plausible that the exchange rate is driven by local and political risk suggesting that it is a partial proxy for a changing political environment (Lim, 2003: 2). The results in Table 8.1. indicate that the impact of unanticipated changes is overwhelmingly negative; $USD\varepsilon_t$ has a negative and statistically significant impact on 15 industrial sectors and the mean coefficient is -0.180. The sole positive and statistically impact is for the mining industry (see Appendix A). Antin (2013: 6) states that the vast majority of South African mining output in the form of minerals is designated for export. This finding, for this sector, supports the hypothesis that exporting firms gain from a depreciation in the value of the domestic currency through increased competiveness, rising demand and sales. However, the strong dominance of a negative relationship between industrial sector returns and $USD\varepsilon_t$ provides support for the general hypothesis that a depreciation is associated with higher input costs that adversely impact returns and that depreciation also potentially reflects heightened political risk.

The impact of unanticipated changes in metal prices, MET_t, is predominantly positive. Returns on 13 industrial sectors are positively and significantly impacted by MET_t and the mean coefficient is 0.155. Partalidou, Kiohos, Giannarakis and Sariannidis (2016: 77; 80) state that metal commodities are often a significant source of export earnings for developing countries and that increases in metal prices are an indicator of economic growth. As South Africa is an emerging market, this suggests that metal prices are also likely to be relevant for the South African stock market. Chen (2010: 127) states that global economic growth has increased the demand for commodities which, in turn, is responsible for rising metal prices suggesting that metal prices are a proxy for global economic conditions. Moolman (2003:294) states that South Africa is a small, open economy that is vulnerable to changes in economic conditions in the rest of the world, implying that changes in metal prices will impact South African stock returns and the South African economy in general through trade channels. Edwards and Alves (2006: 480) support this hypothesis and show that metal exports and related products (iron ingots, aluminium, iron ore, pig iron, etc.) are positioned amongst South Africa's top 20 exports. Therefore, literature indicates that metal prices may be viewed as a proxy for global economic conditions and will impact the South African stock market through trade channels, supporting the observed positive impact.

Innovations in interest rates, LTY_t, as measured by the yield on government bonds with long maturity periods, have a negative and statistically significant impact on returns on 20 industrial sectors. The mean coefficient is -3.920. The impact of interest rates on stock prices has been widely studied in the literature and can be observed directly through the dividend discount model (equation (6.2)). Muradoglu, Taskin and Bigan (2000: 34) propagate a standard argument; the impact of increasing (decreasing) interest rates is to raise (lower) the discount rates used in valuing stocks. This has a negative (positive) impact on returns. Wongbangpo and Sharma (2002: 31) suggest an opportunity cost effect; higher (lower) rates motivate investors to substitute equity holdings for other assets and therefore have a negative (positive) impact on stock prices. It is also suggested that rising (declining) interest rates may negatively (positively) impact financing costs and thereby reduce (increase) profitability. Thorbecke (1997: 638) argues that increasing (decreasing) interest rates impact a firm's net worth and consequently, a firm's ability to invest. The abridged results in Table 8.1. are consistent with the literature.

The final macroeconomic factor, $TLI\varepsilon_t$, is the innovation series of the composite index of leading indicators for South Africa's trading partners. Moolman (2003: 294) argues that given the small size and the open nature of South Africa's economy, South Africa is impacted by changes in economic conditions in the rest of the world and especially those experienced by South Africa's trading partners, the US and developed European economies. It is further suggested by the author that South Africa's vulnerability to external economic conditions has increased during the post 1994 transition and with increasing globalisation and economic interdependence. However, Moolman (2003) finds that this factor is a poor predictor of turning points.[127] Nevertheless, given that $TLI\varepsilon_t$ is significantly and positively associated with returns on 24 industrial sectors and has an overall positive impact, as suggested by a mean coefficient of 2.865, it is possible that this factor has become increasingly important (since Moolman's (2003) study) due to growing international economic interdependence. A Gauteng Provincial Treasury (2013: 19, 30) report on the impact of business cycles on the South African economy provides support for this

[127] Moolman (2003) reports that the highest pseudo $\bar{R}^2$ is for 18 months ahead at 0.096.

hypothesis. South Africa is an open but relatively small economy with its health dependent upon that of its trading partners. It is shown that between 2004 and 2011, the value of South African exports more than doubled and was accompanied by an increase in GDP. However, both GDP and exports decrease between 2008 and 2009 during the global financial crisis, suggesting a link between the health of South African economy and the broader global economy.[128] In conclusion, the report proposes that global economic events impact South Africa through trade channels and that output fluctuations are often driven by external economic events that affect the domestic economy. These arguments suggest that $TLI\varepsilon_t$ impacts stock prices by proxying for and predicting changes in the economic conditions experienced by South Africa's trading partners, which in-turn, are reflected in the domestic business cycle. Positive (negative) changes in the external economic climate will impact the domestic economy through trade channels and indirectly have a propitious (adverse) impact on confidence, affecting expectations of future corporate profitability. This will translate into an impact on stock prices and the hypothesised direction of impact is support by the results.

The present discussion confirms the hypothesised direction of the impact of the factors considered. This suggests that the model specification is sensible and meets *a priori* expectations. All factors appear to be systematic in character, with the exception of BP_{t-1}. Although BP_{t-1} is considered in a number of studies of the South African market, notably those of Moolman (2003) and Szczygielski and Chipeta (2015), its impact appears to be limited. This is in contrast to the results in Table 7.4. which show that this factor is correlated with a substantial number of industrial sectors. It may be that this factor has a weak impact on South African stock returns and its explanatory power is subsumed by the remaining factors that feature in the linear factor model. Nevertheless, this factor is retained, given that it features prominently in literature on the South African stock market and given that it is shown to be a proxy for pervasive influences in the factor score regressions (Table 7.7.).

8.3.2. Residual Market Factors And The Factor Analytic Augmentation

Wei (1988: 888-889) and Van Rensburg (1995: 59) state that the residual market factor beta is non-zero if factors have been omitted from the linear factor model. Insight may be gained

[128] See the Gauteng Provincial Treasury (2013) Quarterly Bulletin, Figure 7.

into factor omission at this stage by giving consideration to the significance of the residual market factors in the benchmark model.

The residual market factor, $M\varepsilon_t$, is orthogonal to the macroeconomic factors in the benchmark model specification. By design, it reflects influences not reflected in the macroeconomic factors, which are required for an adequate description of the underlying factor structure (Chang, 1991: 380). The coefficient on $M\varepsilon_t$ has a positive and statistically significant impact on returns for all industrial sectors and a mean coefficient of 0.664. This is as expected. As reported in Connor (1995: 45), macroeconomic factors by themselves are poor proxies for the underlying pervasive influences in returns. Also, the role of the residual market factor *is* to account for omitted factors in the linear factor model (Berry *et al.,* 1988:31; Section 3.2.; Section 3.4). Therefore, it is expected that this factor will have a significant (and positive) impact on returns. What is not expected, if the residual market factor is an adequate proxy for omitted factors, is that $IM\varepsilon_t$ will have a statistically significant impact on a substantial number of industrial sectors. By construction, $IM\varepsilon_t$ reflects any information not reflected in $M\varepsilon_t$ and the macroeconomic factors. Nevertheless, $IM\varepsilon_t$ is found to have a statistically significant impact on 16 industrial sectors suggesting that the macroeconomic factors and $M\varepsilon_t$ fail to account for unobserved and omitted global factors (Clare & Priestley, 1998: 110). This preliminary evidence indicates that the conventional residual market factor, as applied in this study and in the literature, may be insufficient to resolve underspecification.

What is perhaps most concerning is that the benchmark specification incorporates statistical factors, f_{1t} and f_{2t}, derived from the residuals of equation (8.1) (Section 6.4.1.; equation (6.19)). As suggested by King (1966: 166), the factor analytic augmentation may reflect industry-specific factors as opposed to common factors. The results in Table 8.1. refute this given the widespread significance of these factors. This suggests that these factors represent unobserved common influences as opposed to industry-specific influences. The first factor, f_{1t}, has a significant impact on returns for 20 industrial sectors whereas the second factor, f_{2t}, also has a significant impact on returns for 20 industrial sectors. Returns on 15 industrial sectors are significantly related to *both* factors. Returns on only one sector, the forestry and paper industrial sector, are not explained by either of the two factors (Appendix A). The existence of these factors implies that the assumption of uncorrelated residuals across industrial sectors is violated. By definition, the goal of factor analysis is to derive a set of common factors that account for correlations (Yong & Pierce, 2013: 80). In the macroeconomic APT, these factors can be represented by macroeconomic proxies. The extraction of factors from the residuals of equation (8.1) indicates that common and unobserved factors are still present in the residuals but omitted from the specification. This suggests that the unrestricted specification in equation (8.1) is potentially underspecified even when two residual market factors, $M\varepsilon_t$ and $IM\varepsilon_t$, are used to proxy for omitted factors. This is investigated further in Chapter 10.

8.3.3. Model Assessment

The mean adjusted coefficient of determination, $\bar{R}^2$, reported in Panel B of Table 8.1., is 0.504. The lowest observed $\bar{R}^2$ is 0.171 for the fixed line telecommunications sector and the highest is 0.941 for the mining sector. McElroy and Burmeister (1988: 38) report a somewhat smaller range for their five-factor model with $\bar{R}^2$ ranging between 30% and 50%.[129] The mean $\bar{R}^2$ for Berry *et al.'s* (1988: 38) five-factor model of the 79 industrial sectors in the sample is 0.485 and ranges between 0.15 for the beverages and brewers sector and 0.75 for the office and business equipment industrial sector. These results are closer to those of the benchmark model in this study and the mean $\bar{R}^2$ is comparable.[130] Also reported in Panel B of Table 8.1. are the mean AIC and BIC statistics (Section 6.4.4.). The mean AIC statistic the benchmark specification is -3.348 and ranges between -4.956 for the mining sector and -2.035 for the fixed line telecommunications sector. The mean BIC statistic for the benchmark specification is -3.114 and ranges between -4.736 for the mining sector and -1.797 for the fixed line telecommunications sector. These statistics become more meaningful in Chapter 9 and Chapter 10. The restricted specification and the unrestricted specifications are compared to the benchmark specification and against each other on the basis of the $\bar{R}^2$, AIC and BIC values to determine whether the inclusion of residual market factors improves model specification and provides a more appropriate description of the data and the true return generating process.

[129] $\bar{R}^2$'s for individual regressions are not reported and therefore no mean $\bar{R}^2$ value can be estimated.
[130] The reader is reminded that the results in McElroy and Burmeister (1988) and Berry *et al.* (1988) present a rough benchmark for comparison. This is because the model, the sample, the sample period and the purpose of these studies differs from those of this study.

Finally, the absolute mean differences between ML and least squares coefficient estimates for the benchmark specification are reported in Panel A of Table 8.1. The absolute differences range from 0.0006 for the intercept and 0.221 for $TLI\varepsilon_t$. Differences are statistically significant for two factors; MET_t and $TLI\varepsilon_t$ although the Wilcoxon matched-pairs signed-ranked test suggests that differences are also statistically significant for $USD\varepsilon_t$ but contradicts the result of the *t*-test for MET_t. If the model is adequately specified, differences will be attributable to the presence of pure heteroscedasticity.[131]

Differences between ML and least squares coefficient estimates can be seen as a measure of bias induced by underspecification by permitting impure heteroscedasticity to impact model parameters (Bera *et al.,* 1988).[132] What is of interest is whether these differences increase in magnitude for the restricted and unrestricted specifications. Factor omission will impact the structure of the conditional variance and this is hypothesised to impact coefficient estimates in the linear factor model. As these latter specifications exclude the factor analytic augmentation, and the residual market factors in the restricted specification, any differences can be attributed to impure heteroscedasticity. Much like the $\bar{R}^2$ and AIC and BIC statistics, differences between ML and least squares coefficient estimates permit for comparisons between specifications and are further considered in Chapter 9 and Chapter 10.

8.4. MODEL DIAGNOSTICS AND ROBUSTNESS

Table 8.2. reports the abridged results of the model diagnostic tests outlined in Section 6.4.5.

[131] The significance of the mean absolute differences for MET_t and $TLI\varepsilon_t$ between the ML and least squares coefficient estimates may be attributed to departures from normality of the residuals. Under an assumption of normality, least squares and ML estimators will be identical (Wooldridge, 2013: 815). Departures from normality may themselves be driven by pure and impure heteroscedasticity, which is reflected in the return distribution and suggested by leptokurtosis (Akgiray, 1989: 62).

[132] For the purposes of consistency, all comparisons are against least squares coefficients of the benchmark specification which do not reflect the conditional variance structure and are assumed to be BLUE. Any difference in the coefficients of the subject model (the restricted specification in Chapter 9 and the unrestricted models in Chapter 10) and the benchmark specification can be attributed to the structure of the conditional variance, which will be affected by underspecification. It is therefore anticipated that the absolute differences between ML and the least squares coefficients will increase with underspecification and will reflect departures from the BLUE properties. The estimated coefficients will reflect impure conditional heteroscedasticity. As argued by Bera *et al.* (1988), the greater the level of conditional heteroscedasticity which is dependent upon factors included and excluded from a specification, the greater the difference between estimated ML and least squares coefficients.

Table 8.2: Abridged Benchmark Model Diagnostics

Test	Mean Value	Total Sig.
F-Test	36.528	26/26
JB Test	11.511	15/26
$Q(1)$	1.744	5/26
$Q(5)$	6.586	4/26
$Q^2(1)$	0.236	0/26
$Q^2(5)$	3.812	0/26
ARCH(1)	0.235	0/26
ARCH(5)	0.786	0/26

Notes:
Significance is recorded at the 10% level of significance. *F*-Test reports the results for Wald's test of linear restrictions jointly equating all explanatory factors in the respective specifications to zero. JB Test summarises the results of the Jarque-Bera test for normality. $Q(1)$ and $Q(5)$ are Ljung-Box *Q*-statistics indicating whether serial correlation in the residuals is statistically significant at the first order and jointly up to five orders of serial correlation respectively. $Q^2(1)$ and $Q^2(5)$ are Ljung-Box test statistics for non-linear dependence in the residuals at the first order and jointly up to five orders. ARCH(1) and ARCH(5) are Lagrange Multiplier (LM) tests for ARCH effects in the residuals at the first and fifth orders respectively. Mean Value reports the mean of the respective test statistics and Total Sig. reports the number of instances in which the results of the respective tests applied are statistically significant.

As expected, the results of the *F*-test indicate that the macroeconomic factors, the residual market factors and the factor scores are jointly statistically significant for all industrial sectors. This confirms the overall significance of the benchmark linear factor model proposed to describe South African stock returns (Sadorsky, 2001: 25). The JB test statistic is statistically significant for the residuals of 15 of the 26 industrial sectors suggesting that the majority of sectors exhibit conditional non-normality. This confirms the appropriateness of using QML estimation with Bollerslev-Wooldridge robust standard errors to estimate specifications for which the residuals depart from normality. Comparisons of the number of instances of residual normality or lack thereof for the restricted and unrestricted specifications and the benchmark specification are reported in Chapter 9 and Chapter 10.

Q-statistics indicate that the model residuals are generally free of serial correlation at the first and up to the fifth orders with the exception of a limited number of sectors. These are the chemicals, industrial engineering, fixed line telecommunications, non-life insurance, life insurance and equity investment and instruments sectors – a total of six sectors (at both first and/or up to the fifth order of serial correlation, see Table A1.1. in Appendix A). For sectors that exhibit statistically significant serial correlation, the null hypothesis of omitted non-linear combinations of factors for the RESET test is not rejected for only a single industrial sector, the industrial engineering sector. This provides general support for the assumption of the linearity of the return generating process underlying the South African stock market.

Although residual serial correlation may imply that the benchmark model may have been incorrectly specified, the results, supported by the application of the RESET test, imply that observed residual serial correlation is pure in nature (Studenmund, 2014: 325). *Q*-statistics for the squared residual series at the first and fifth orders indicate an absence of non-linear serial correlation. The absence of ARCH effects in the residuals is confirmed by statistically insignificant ARCH(1) and ARCH(5) LM test statistics (Akgiray, 1989; Engle, 2001: 162). This implies that any remaining efficiency loss in coefficient estimates is not associated with uncaptured heteroscedasticity or underspecified ARCH*(p)* or GARCH*(p,q)* models used to model the conditional variance.

The (unreported) results[133] of the least squares estimation with HAC standard errors and MM estimation for the benchmark model across sectors are, with very few exceptions, consistent in terms of coefficient size and direction of impact with the results obtained from ML estimation. A limited number of coefficients that are statistically significant in individual regressions are no longer statistically significant following least squares estimation and MM estimation and *vice versa*. A closer inspection of the individual inconsistencies implies weak significance or weak insignificance in the first instance.[134] For example, least squares results with HAC standard errors and those of MM estimation for the media sector indicate that BP_{t-1} is statistically significant at the 10% and 5% levels of significance (*p*-values of 0.086 and 0.032) respectively, whereas the associated coefficient is (weakly) statistically insignificant if estimated using ML estimation (*p*-value of 0.108). Another example is the electronic and electrical equipment industrial sector. The results of least squares estimation with HAC standard errors are consistent with those of ML estimation; $LEAD_{t-1}$ is statistically insignificant. In contrast, the results of MM estimation indicate that $LEAD_{t-1}$ is (weakly) statistically significant (with a *p*-value of 0.084).

In total, there are 14 inconsistencies between the ML and least squares estimates and 16 inconsistencies between ML and MM estimates. However, six of the 14 inconsistencies for the least squares methodology and 10 of the 16 inconsistences for MM estimation are somewhat ambiguous. In such cases, coefficients are either weakly statistically significant (insignificant) following least squares and MM estimation but insignificant (significant)

133 The results of this analysis are available upon request.

134 The definition of what comprises a weakly insignificant or significant coefficient is somewhat arbitrary. This study defines a weakly insignificant coefficient as having a *p*-value of between 0.1 and 0.14 and a weakly significant coefficient as having an approximate *p*-value of 0.085.

following ML estimation. As for the remaining unambiguous inconsistencies, some inconsistencies are to be expected (for example see Andersen *et al.,* 2003: 48). Unexplained and large inconsistences in significance (and the associated *p*-values) may be attributed to advantages and drawbacks of the respective estimators and not the validity of the benchmark specification. For example, while ML estimation with ARCH/GARCH errors may be well-suited to modelling volatility dynamics, non-linear dependence and excess kurtosis, the ARCH(*p*) and GARCH(*p,q*) models of conditional variance may be themselves misspecified[135] (Elyasiani & Mansur, 1998: 548; Engle & Patton, 2007; Andersen *et al.,* 2003: 46; Lee, 2011: 757). This is a limitation of the ARCH and GARCH methodology, namely the difficulty associated with identifying the most appropriate ARCH(*p*) or GARCH(*p,q*) model to describe the conditional variance structure of a given series. On the other hand, while the least squares methodology with HAC standard errors is robust to serial correlation and heteroscedasticity of unknown form, robust regression techniques, such as MM estimation, have been developed to provide robust estimates in the presence of influential observations and outliers that may impact coefficient estimates (Andersen, 2008; Wooldridge, 2013: 432).

In conclusion, the benchmark model estimated using ML estimation with ARCH(*p*) or GARCH(*p,q*), errors appears to yield theoretically (Section 8.3.1.) as well as empirically robust and reliable results. Significant *F*-tests confirm the overall significance of the benchmark model consisting of macroeconomic factors, the two residual market factors and a factor analytic augmentation across industrial sectors. Residuals are generally serially uncorrelated and there is no evidence of non-linear dependence or ARCH effects.

8.5. VARIANCE AND CONDITIONAL VARIANCE

The results in Table 8.3. indicate that the mean residual variance, $\sigma^2_{BM\varepsilon_i}$, is 0.002483. The sector with the largest residual variance is the software and computer services sector with a residual variance of 0.00682. The sector with the smallest residual variance is the mining sector with a residual variance of 0.000368. In the latter case, this is the sector with the highest $\bar{R}^2$ of 0.941. This result may be understood by decomposing variance into systematic and sector-specific components. The decomposition of variance implies that variation in the dependent factor (sector returns) arises from regression variance (the

[135] There is a debate as to whether any other ARCH/GARCH-type specification is superior to a GARCH(1,1) specification. See Hansen and Lunde (2005) and Awartani and Corradi (2005) for a discussion of the performance of various ARCH and GARCH specifications.

model, quantified by the $\bar{R}^2$) attributable to the explanatory factors *and* residual variance. The larger the regression variance, the smaller the residual variance (Greene, 2012: 1007). In an underspecified model, residual variance will be higher and the regression variance will be lower because both omitted systematic factors and idiosyncratic factors will be relegated to the residuals (Lehmann,1990: 72). As the benchmark model explains over 90% of the variance in returns on the mining sector, the residual variance is small. This relationship also approximately holds for the software and computer services sector which has the highest residual variance and fourth lowest $\bar{R}^2$ of 0.314. The same may be said of the other sectors that have a low $\bar{R}^2$ (lower than for the software and computer services sector), but high residual variance. The fixed line telecommunications sector has an $\bar{R}^2$ of 0.171 and residual variance of 0.00665, the automobile and parts sector has an $\bar{R}^2$ of 0.205 and a residual variance of 0.006090, the pharmaceuticals sector has an $\bar{R}^2$ of 0.284 and residual variance of 0.003567 (Table A1.1. in Appendix A). This suggests that as the residual variance decreases (increases), the $\bar{R}^2$ increases (decreases) (see footnote).[136] A comparison of the mean residual variance across specifications combined with the Brown-Forsythe test of the equality of residual variance across specifications for each sector becomes both informative and relevant in quantifying the impact of factor omission on residual variance and importantly, coefficient standard errors and (Section 6.4.6.). This is explored in Section 9.5. and Section 10.5.

[136] A correlation coefficient of -0.656 between residual variance values and the $\bar{R}^2$ values confirms this.

Table 8.3: Benchmark Model Residual Variance And Conditional Variance Structure

Panel A: Residual Variance			
	Mean Value	**Minimum**	**Maximum**
$\sigma^2_{BM\varepsilon_i}$	0.002483	0.000368 Mining	0.00682 Software & computer serv.
Panel B: Conditional Variance Structure			
Model	**ARCH(1)**	**GARCH(1,1)**	
Frequency	18	8	
	Mean Coeff.	**Mean Coeff.**	
ω	0.002 (17)	0.0001 (0)	
α_1	0.103 (3)	0.103 (4)	
β_1		0.833 (8)	
Sig. *F*-Test	3/18	8/8	
Panel C: Conditional Heteroscedasticity			
α_i	0.103		

Notes:
The asterisks, ***, ** and *, indicate statistical significance at the respective 1%, 5% and 10% levels of significance.
In Panel A, the Mean Value is the mean of the residual variance across sectors. The Minimum and Maximum values are the lowest and highest residual variances observed for the respective sectors.
In Panel B, Frequency is indicative of the number of instances of each ARCH(*p*) or GARCH(*p,q*) model applied. Negative ARCH coefficients, which are indicative of the absence of conditional heteroscedasticity and are therefore statistically insignificant, are rounded to zero in aggregation. This is consistent with the non-negativity constraint ($\alpha_i \geq 0$) placed upon the ARCH coefficient in the ARCH(*p*) specification (Poon, 2005: 38). The numbers in brackets ( ) next to each mean value indicate the number of statistically significant coefficients for each ARCH(*p*) or GARCH(*p,q*) specification at the 10% level of significance. Sig. *F*-Test reports the number of significant instances of Wald's test of linear restrictions for the ARCH and GARCH coefficients. The null hypothesis is that ARCH and GARCH coefficients are jointly equal to zero. In Panel C, α_i is the arithmetic mean of the ARCH coefficients across sectors.

The results in Panel B of Table 8.3. indicate that the conditional variance structures of 18 industrial sectors are described by the (short-memory) ARCH(1) process (Elyasiani & Mansur, 1998: 541). Only three industrial sectors to which the ARCH(1) model is applied report a statistically significant *F*-statistic. These are the industrial transport, automobile and parts and the travel and leisure industrial sectors. This indicates that the variance underlying these sectors is of a time-varying nature but not for the other sectors for which the *F*-statistic is insignificant. The conditional variance of eight industrial sectors is modelled as a GARCH(1,1) process and the *F*-statistic is statistically significant for all eight GARCH(1,1) specifications. A preliminary analysis (unreported) shows that this specification is appropriate for series that exhibit higher order non-linear residual dependence and higher order ARCH effects. For example, non-linear dependence in the squared residuals is observed up to the 20th order and ARCH effects are observed above the 10th order for returns on the chemicals sector after an ARCH(1) model is fitted. This is no longer the case when a GARCH(1,1) model is fitted to this series. Taken in their entirety, these results indicate that most (15 out of 26) residual series derived from the benchmark model are not characterised by time-varying variance and there is an absence of non-linear dependence and ARCH effects in these residual series.

Finally, the mean conditional heteroscedasticity parameter α_i, which has a value of 0.103, is reported in Panel C of Table 8.3. This value is also reported in the subsequent chapters for the restricted and unrestricted models. Comparisons are made on the basis of this value in the chapters that follow, to establish how conditional heteroscedasticity is impacted by factor omission.

8.6. PREDICTIVE ABILITY

To gain insight into the ability of the benchmark model to replicate (fit/predict) actual returns accurately, the analysis begins with a consideration of the aggregate residuals, ε_{it}, derived from the benchmark model in equation (8.2). The results in Panel A of Table 8.4. indicate that overall, prediction errors do not differ significantly from zero. Results are consistent across the *t*-test and the Wilcoxon test. These results favour the benchmark specification (Chang, 1991: 387). It remains to be seen whether the restricted model also produces such favourable results and whether the unrestricted market model approximates these results. This is investigated in Section 9.6. and Section 10.6.

Table 8.4: Summary Of Mean Errors And Theil's *U* Statistic For The Benchmark Model

Panel A: Mean Errors			
	Mean Value		
ε_{it}	-0.0005692		
Panel B: Theil's *U* Statistic And Decomposition			
	Mean Value	**Minimum**	**Maximum**
Theil *U*	0.395	0.120 Mining	0.602 Fixed line telecom.
Bias (U_{BIAS})	0.001036	0.000000 Industrial metals & mining	0.017506 Software & computer services
Variance (U_{VAR})	0.178215	0.022184 Mining	0.509499 Software & computer services
Covariance (U_{COV})	0.820820	0.472995 Software & computer services	0.976984 Mining

Notes:
The asterisks, ***, ** and *, indicate statistical significance at the respective 1%, 5% and 10% levels of significance. In Panel A, the Mean Value is the respective arithmetic mean of the residual terms. A paired-sample *t*-test is applied to test the null hypothesis that the mean value of the residuals differs significantly from zero. In Panel B, the Mean Value is the arithmetic mean of the respective measures of predictive accuracy. The Minimum and Maximum are the minimum and maximum values associated with the respective measures of accuracy for the respective sectors. A superscript "W" indicates a discrepancy between the results of the *t*-test and the Wilcoxon test.

The mean Theil U statistic and the bias, variance and co-variance proportions for the benchmark model are reported in Panel B of Table 8.4. The mean U statistic is 0.395 and ranges between 0.120 for the mining sector and 0.602 for the fixed line telecommunications sector. Frank (2009: 58) states that the U statistic is analogous to the $\bar{R}^2$ and that a large value indicates poor model performance. Although a U statistic of zero is desirable, these results show that predictive performance varies across sectors and that the benchmark model does not replicate the actual return series with perfect accuracy across industrial sectors. This raises a somewhat pertinent but potentially complex question (and beyond the scope of this study) as to what is an acceptable or "good" Theil U statistic or analogously $\bar{R}^2$ for an APT-type linear factor model.

The mean bias proportion, U_{BIAS}, is 0.001036[137] and ranges between zero for the industrial metals and mining sector and 0.017506 for the software and computer services sector. Both the range of values and the mean bias proportion are indicative of low values, close to zero, suggesting that there is almost no bias in the predicted values across industrial sectors. The bias proportion can be viewed as an additional measure of the overall bias in the estimated model coefficients; if coefficients are biased, the predictions will be biased (Walther & Moore, 2005: 816). In the econometric framework applied in this study, coefficient bias will be a consequence of factor omission that impacts the structure of the conditional variance that enters the log-likelihood function (Bera *et al.,* 1988: 212; consequence 1) in Section 5.3.1.). Therefore, the bias proportion will be useful in establishing the impact of factor omission on parameter bias in Chapter 9 and Chapter 10 as well as useful in establishing the direct impact of factor omission on predictive accuracy.

The mean variance proportion, U_{VAR}, is 0.178215 and ranges between 0.022184 for the mining sector and 0.509499 for the software and computer services sector. The ability of a model to replicate the variance of the actual series will depend on the ability of the model to accurately replicate the magnitude of the actual observations. The mean covariance proportion, U_{COV}, is 0.820820 suggesting that although most of the prediction error is attributable to unsystematic or residual components, some prediction error is systematic in nature (Brooks, 2008: 258). As with the bias and variance proportions, the mining sector has a desirable covariance proportion of 0.976984 , which is close to unity. The software and

[137] Brooks and Tsolacos (2010: 272) suggest that a value of above 0.1 or 0.2 for the bias proportion is concerning. None of the series exhibit bias proportions of this magnitude

computer services industrial sector has the least desirable (lowest) covariance proportion of 0.472995.

The *U* statistics and the bias, variance and covariance proportions obtained from the restricted and unrestricted specifications in Chapter 9 and Chapter 10 respectively are compared against those reported in Table 8.4. This is to establish whether factor omission results in inferior predictive performance and whether the inclusion of the residual market factors translates into performance, which is comparable to that of the benchmark model by reducing the bias and variance proportions and maximising the covariance proportions for the unrestricted models.

8.7. FACTOR OMISSION

To explore the structure of the residual correlation matrix, the MAP test is applied to derive the optimal factor solution (Section 6.4.8.).[138] It follows that if there are no common factors in the residuals, the optimal factor solution will yield no factors. The results are summarised in Panel A for the full period and in Panel B of Table 8.5. for the subperiods.

Table 8.5: Summary Of Factor Analysis Of Benchmark Model Residuals

Panel A: Full Period Factor Analysis		
Factors extracted	**Mean Communality**	**Mean Uniqueness**
1	0.066	0.934
Panel B: Subperiod Factor Analysis		
Period: 2001M016 to 2008M12		
0	**-**	-
Period: 2009M01 to 2016M12		
1	0.097	0.903

Notes:
Mean Communality is the mean proportion of common variance explained across return series by the statistical factors extracted on the basis of the MAP test. Mean Uniqueness is the mean proportion of variance across return series attributable to the return series themselves and not to systematic factors**.**

For the entire sample period, a single factor is extracted from the residual correlation matrix and accounts for 6.6% of the variation in the residuals (mean communality). A closer examination of the results of the factor analysis is revealing. While the mean communality, which indicates the proportion of shared variance explained by this factor, is 0.066 (as above), the mean uniqueness, which is indicative of variance that is specific to each sector, is 0.934. In other words, while 6.6% of the variance of the residuals may be explained by a potentially omitted common factor which may be non-trivial, 93.4% of variance is attributable to unique factors in the residual series (Walker & Madden, 2008: 326). Furthermore, a

[138] The scree test is inconclusive as there is no distinct flexion point.

closer examination of the 26 sectors reveals that only five sectors have communalities that are greater than 0.15 suggesting that this factor explains a substantial proportion of residual variance for this specific subset of industrial sectors. These are the chemicals, mining, beverages, general retailers and banks industrial sectors.[139] The sector with the highest loading on the single extracted factor is the general retailers sector with a loading of -0.492 and a corresponding communality of 0.242. When the residual series for the general retailers sectors is excluded from the residual factor analysis and the MAP test is applied, no factors are extracted.[140] This implies that the single factor that is extracted may be the result of strong interdependence between a limited number of industrial sectors. Such a factor will meet the definition of a pseudofactor; a factor that explains variance for a limited number of series but has little explanatory power across all series in the sample (Connor, 1995: 44).

For the subperiod analysis reported in Panel B of Table 8.5., the MAP test fails to identify a single factor for the 2001M016 to 2008M12 period but results in the extraction of a single factor for the 2009M01 to 2016M12 period.[141] For the 2009M01 to 2016M12 period, this factor is associated with a mean communality 0.097 whereas the mean uniqueness is 0.903. The mean communality and the uniqueness values are higher and (slightly) lower respectively than those reported for the entire sample period. The subperiod analysis suggests that the factor extracted is a transitory factor, which meets a further definition of a pseudofactor, namely a factor that is important for specific time periods (Kryzanowski & To, 1983: 42).

Meyers (1973: 698) argues that if a (linear factor-type) model is valid, then no factors should be reflected in the residual dependence structure. However, if factors remain, but represent transitory statistical artefacts, then the validity of the model in question need not be interrogated. The factor analysis conducted in this section suggests that these factors are indeed such statistical artefacts. Firstly, it appears that the extracted factor is specific to a subset of industrial sectors and is therefore a pseudofactor. Secondly, this factor seems to emerge during the second half of the sample period and appears to be transitory in nature.

[139] Detailed results of the factor analysis are available upon request.

[140] The software used in the analysis is Eviews 9.5. Eviews reports a communality of 0 and a uniqueness of 1 for all industrial sector residual series when the general retailers sector is excluded and the MAP test is applied. In other words, no common factors are identified.

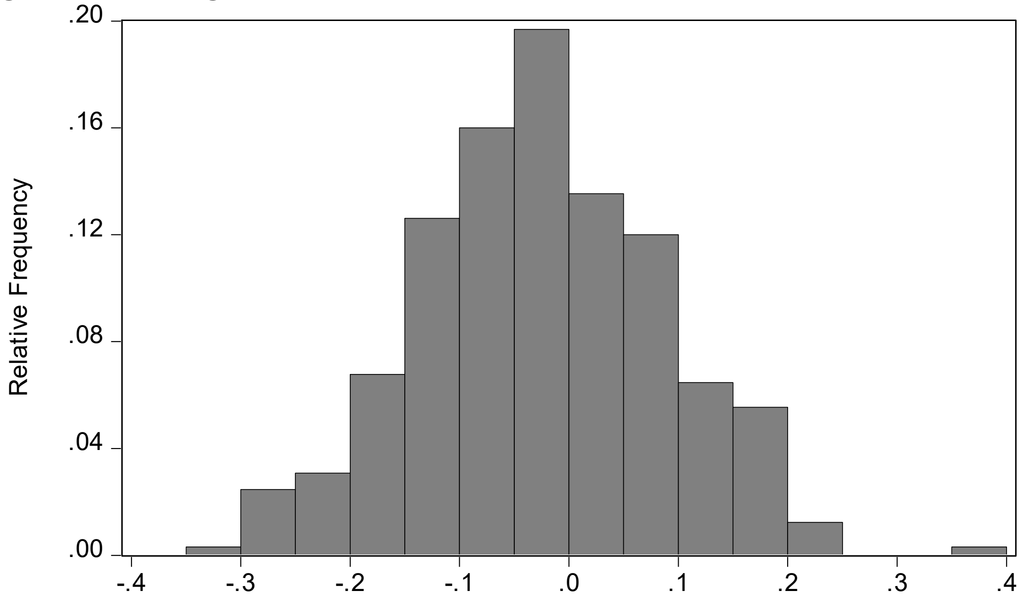
[141] It is appropriate to consider that the MAP test derives the number of factors that will result in the residual correlation matrix closely resembling an identity matrix (Section 6.3.1). If the MAP test fails to identify additional factors, then the addition of any additional extracted factors (through respecification) will not induce the resultant residual correlation matrix to resemble an identity matrix more closely.

Finally, the existence of this factor does not appear to alter the general results of the benchmark model.[142] In summary, these results indicate that the benchmark model represents a well-formulated specification against which the restricted and unrestricted versions of the model may be compared to investigate the impact of factor omission.

8.8. THE RESIDUAL CORRELATION MATRIX

Figure 8.1. presents a histogram of the ordinary correlation coefficients of industrial sector residuals from the benchmark specification. Table 8.6. summarises the distribution of correlation coefficients.

Figure 8.1: Histogram Of Benchmark Model Residual Correlation Coefficients



[142] As a final test of the robustness of the results, the single factor that is extracted for the full sample period is incorporated into the benchmark specification for the five sectors that have communalities of above 0.15. With the exception of the banks industrial sector, there are no changes in the significance or magnitudes of the coefficient estimates for the other four sectors. For the banks industrial sector, the coefficient on MET_t is now statistically significant and there are some noticeable changes in the magnitudes of the coefficients on the other factors. As expected, with the exception of the mining sector, the $\bar{R}^2$ increases somewhat for each sector. Nevertheless, the overall results indicate that the existence of possible pseudofactors does not invalidate the results of the benchmark model nor is the existence of pseudofactor likely to have a widespread impact on the results of the benchmark model as the remaining communalities are below 0.15.

Table 8.6. Distribution Summary Of The Benchmark Model Residual Correlation Matrix

Bin	Frequency	Relative Frequency	Cumulative Frequency
$-0.5 < \rho_{ij} \leq -0.4$	0	0.000%	0.000%
$-0.4 < \rho_{ij} \leq -0.3$	1	0.308%	0.308%
$-0.3 < \rho_{ij} \leq -0.2$	18	5.538%	5.846%
$-0.2 < \rho_{ij} \leq -0.1$	63	19.385%	25.231%
$-0.1 < \rho_{ij} \leq 0$	116	35.692%	60.923%
$0 < \rho_{ij} \leq 0.1$	83	25.538%	86.462%
$0.1 < \rho_{ij} \leq 0.2$	39	12.00%	98.462%
$0.2 < \rho_{ij} \leq 0.3$	4	1.231%	99.692%
$0.3 < \rho_{ij} \leq 0.4$	1	0.308%	100.000%
$0.4 < \rho_{ij} \leq 0.5$	0	0.000%	100.000%
Total	**325**	**100%**	**100.000%**
Mean	-0.024***		
Minimum	-0.320		
Maximum	0.396		

Notes:
The asterisks, ***, ** and *, indicate statistical significance at the respective 1%, 5% and 10% levels of significance. The *t*-test is applied to test the hypothesis that the mean of the correlation coefficients does not differ significantly from zero. The Wilcoxon matched-pairs signed-rank test is applied as a confirmatory test and the superscript "W" indicates that the Wilcoxon matched-pairs signed-rank test contradicts the results of the paired-sample *t*-test. . Bin represents ranges of correlation coefficients and Frequency reports the number of correlation coefficients that fall within each range. Relative Frequency is the percentage of correlation coefficients that fall within the respective ranges. Cumulative Frequency is the running total of all previous frequencies in percentage terms. Mean is the mean value of the correlation coefficients in the correlation matrix and the Minimum and Maximum are the lowest and largest correlation coefficients observed.

As indicated by Figure 8.1, the majority (61.230%) of (off-diagonal) residual correlation coefficients lie within the -0.1 to 0.1 range. This corresponds to a total of 199 correlation coefficients. Importantly, 71.692% (233 out of 325) of correlation coefficients fall within the -0.12 and 0.12 range. Coefficients that are greater than approximately 0.12 in (absolute) magnitude tend to be statistically significant at the 10% level of significance. This indicates that the majority of residual correlation coefficients in the residual matrix are statistically insignificant. However, a substantial number of correlation coefficients, 28.308% (92 out of 325), are greater than 0.12 in absolute terms.

The mean level of correlation is -0.024, which is somewhat encouraging, given that this represents an almost non-existent level of correlation although the *t*-test indicates that this level of correlation is significantly different from zero. This is in stark contrast to the higher mean correlation of 0.375 for the actual return series (Table 7.1.). Correlation coefficients range between -0.320 and 0.396 although these observations appear to be outliers as suggested by the summary in Table 8.6. This again differs from the correlation coefficients of the actual return series, which range between 0.048 and 0.673 and represents a downward shift in the extreme values (Table 7.1.). The large proportion of negative correlation coefficients (60.923%, 198 out of 325) in the residual correlation matrix of the benchmark model contrasts with the all positive correlation coefficients in the actual return correlation matrix. This is indicative of substantially different correlation structure to that of the actual return series. Such a change in the correlation structure is expected if the factors that feature in the specification capture the co-movement attributable to pervasive influences in returns.

Table 8.7. reports the correlation matrix of the benchmark residuals. Of the 32 positive correlation coefficients in Table 8.7., 28 correlation coefficients are found outside of the economic sector submatrices (denoted by the dashed lines). Ostensibly, this is a somewhat concerning finding as it suggests that there may be a pervasive factor or set of factors that is omitted from the benchmark model. However, the factor analysis conducted in Section 8.7. suggests that such interdependence may be limited to a few sectors or/and is driven by transient components and does not invalidate the benchmark specification. The absence of a substantial number of positive pairwise correlations for industrial sectors within the economic sector submatrices suggests an absence of economic sector-specific factors. Of the 60 negative correlation coefficients, 12 coefficients are within the economic sector submatrices whereas the remainder is outside of the economic sector submatrices. According to King (1966: 153), such a finding points towards the presence of inter-economic sector factors.[143] Any such factors, if they exist, are not systematic. As the observed correlations do not appear to follow any specific pattern, unlike the structure of the return series discussed in Section 7.2., which are overwhelmingly positive and significant, it can be argued together with the results in Section 8.7. that the residual correlation observed in Table 8.7. does not pose a challenge to the validly of the benchmark model.

[143] The present analysis of the residual correlation matrix is motivated by that of King (1966), who applies a somewhat different criterion to analyse residual covariance. Defining an absolute residual covariance of 0.00100 as arbitrarily "large", King (1966: 153) reports that there are 165 large covariances out of 1953 estimated covariances. It is partially on the basis of this analysis that the author concludes that the market model fails to account for industrial factors in the residuals. For comparative purposes and as robustness check, this study also considers residual covariance and relies upon King's (1966) definition of a large covariance. Out of the 325 estimated covariances, none may be classified as large. If King's (1966) definition of a large covariance (analogously correlation) is adopted, the assumption of $E(\varepsilon_{it}, \varepsilon_{it}) = 0$ holds in this study.

Table 8.7: Correlation Matrix Of Benchmark Model Residuals

	J135	J173	J175	J177	J235	J272	J273	J275	J277	J279	J335	J353	J357	J453	J457	J533	J537	J555	J575	J653	J835	J853	J857	J877	J898	J953
J135	1.000																									
J173	0.117	1.000																								
J175	-0.193	-0.094	1.000																							
J177	-0.115	-0.174	-0.226	1.000																						
J235	0.020	0.091	-0.067	-0.265	1.000																					
J272	-0.136	0.008	0.142	0.087	0.099	1.000																				
J273	0.075	0.076	-0.178	-0.116	0.103	-0.025	1.000																			
J275	0.042	0.140	-0.204	-0.234	-0.087	-0.104	0.191	1.000																		
J277	0.003	0.081	-0.253	-0.050	-0.271	-0.206	-0.197	0.041	1.000																	
J279	-0.036	0.192	-0.046	0.035	0.091	0.037	-0.056	-0.140	-0.046	1.000																
J335	0.160	0.103	-0.017	-0.061	-0.192	-0.071	-0.022	0.054	-0.037	-0.038	1.000															
J353	0.111	0.148	0.079	-0.130	-0.039	0.022	0.197	0.187	-0.001	-0.104	0.096	1.000														
J357	-0.112	-0.083	0.012	0.050	-0.190	-0.032	-0.020	-0.128	-0.117	-0.101	-0.039	0.133	1.000													
J453	0.050	0.092	-0.122	-0.045	-0.071	-0.020	-0.049	-0.035	-0.042	-0.122	-0.074	0.156	-0.089	1.000												
J457	-0.032	-0.053	0.063	0.068	-0.110	-0.201	-0.019	-0.080	-0.191	-0.093	0.044	0.006	-0.023	0.109	1.000											
J533	-0.166	-0.090	0.084	0.157	-0.102	0.096	-0.244	-0.168	-0.108	-0.158	-0.107	-0.073	0.047	-0.042	-0.009	1.000										
J537	-0.320	-0.139	0.009	0.224	-0.217	-0.078	-0.146	-0.128	-0.083	-0.109	-0.015	-0.227	-0.050	-0.116	0.088	0.153	1.000									
J555	0.102	-0.101	0.018	0.153	-0.032	-0.017	-0.033	0.008	0.170	0.074	0.011	-0.089	-0.140	0.011	-0.016	-0.160	-0.068	1.000								
J575	0.040	0.085	-0.137	-0.065	-0.009	-0.109	-0.144	0.128	0.040	0.074	-0.017	0.148	-0.023	0.029	-0.087	-0.248	-0.273	0.029	1.000							
J653	0.008	-0.158	-0.099	0.204	-0.102	-0.044	-0.051	-0.081	0.025	0.000	0.019	-0.260	-0.009	-0.113	0.087	-0.020	0.046	0.158	-0.029	1.000						
J835	-0.240	-0.158	0.242	0.141	-0.026	0.054	-0.147	-0.020	-0.069	-0.132	-0.157	-0.278	-0.087	-0.101	-0.008	-0.016	0.070	-0.194	-0.107	-0.047	1.000					
J853	-0.056	-0.071	0.008	-0.074	-0.132	-0.027	-0.065	-0.094	0.032	-0.189	-0.049	0.100	0.079	0.075	-0.047	0.047	-0.173	-0.140	-0.053	-0.019	-0.049	1.000				
J857	0.030	0.000	0.214	0.087	-0.085	-0.153	-0.067	0.142	-0.018	-0.090	0.002	-0.124	-0.005	0.057	0.016	-0.013	-0.172	-0.173	-0.045	-0.108	0.190	0.112	1.000			
J877	0.102	-0.096	0.077	0.396	0.076	-0.057	0.157	-0.068	-0.121	-0.021	-0.013	-0.058	-0.055	-0.256	-0.085	-0.013	-0.068	-0.039	0.014	0.027	0.086	-0.142	-0.044	1.000		
J898	0.118	0.098	0.009	0.150	0.060	-0.071	0.019	-0.046	-0.007	0.015	0.115	0.199	0.017	-0.020	0.017	-0.078	-0.164	-0.023	0.081	-0.026	-0.254	0.095	0.072	-0.020	1.000	
J953	-0.045	-0.063	0.099	0.172	-0.087	-0.112	0.037	-0.071	0.082	-0.017	0.076	-0.189	-0.050	-0.087	0.048	-0.071	-0.110	0.182	0.090	0.144	-0.064	0.162	0.106	0.103	0.008	1.000

Table 8.8. summarises the results of Jennrich's (1970) test of the equality of correlation matrices.

Table 8.8: Tests Of Matrix Equality For The Benchmark Model

Hypothesis	χ^2 Statistic	Reject
$B_{26} = A_{26}$	958.629***	Reject
$B_{26} = I_{26}$	826.360***	Reject

Notes:
The asterisks, ***, ** and *, indicate statistical significance at the respective 1%, 5% and 10% levels of significance. Hypothesis is the hypothesis that is being tested relating to the equality of two matrices. χ^2 Statistic is the resultant test statistic of the Jennrich test and Reject indicates whether the null hypothesis of equality between two matrices is rejected. B_{26} denotes the residual correlation matrix derived from the benchmark model. A_{26} denotes the residual correlation matrix of the actual return series. I_{26} denotes the identity matrix.

A formal comparison of the equality of the resultant benchmark model residual correlation matrix against alternatives begins with a test of the equality of the benchmark model residual correlation matrix, B_{26}, and the matrix of actual returns, A_{26}. The Jennrich (1970) test is applied and the null hypothesis of the equality of B_{26} and the correlation matrix of the actual return series, A_{26}, is rejected. This supports the preceding results in this chapter (specifically those in Section 8.3. and Section 8.7.) and points towards the ability of the benchmark model to account for common sources of co-movements in returns. Next, the equality of the benchmark model residual correlation matrix, B_{26}, and the identity matrix, I_{26}, is tested. The Jennrich χ^2 statistic is 826.360 and the null hypothesis of equality is rejected. This implies that the assumption that residuals are uncorrelated across industrial sectors does not hold. The off-diagonal diagonal elements of the residual matrix are in a number of instances and as evident from Table 8.7., non-zero.

The results presented in this section and the preceding section indicate that while there may be some remaining pairwise correlation in the residuals and that the residual correlation matrix does not resemble an identity matrix, correlation is unlikely to be driven by the omission of important systematic factors. The remaining significant correlations in Table 8.7. appear to be sporadic and do not follow a pattern. Moreover, the assumption of uncorrelated residuals may in itself be excessively restrictive and unattainable (Connor & Korajczyk, 1993: 1264). While theoretically desirable, such an outcome may not be achievable in practice. The APT requires that systematic factors are incorporated into the linear factor model (Spyridis *et al.,* 2012: 44). This does not preclude the existence of factors that are individually trivial (and therefore do not have a significant impact on the estimation of the linear factor model) or that are non-general in nature in that they are important for specific subsets of assets or are important for specific time periods (transitory) (Kryzanowski & To, 1983: 42). Factors of this nature may be broadly termed pseudofactors and may be responsible for the low and isolated instance of statistically significant residual correlation observed in Table 8.7. The existence of such factors is unlikely to invalidate the empirical content of the linear factor model and the APT (Beenstock & Chan, 1986: 129). If the existence of these factors is responsible for a departure from the assumption of uncorrelated residuals across series in the correlation matrix but has no broad impact on estimation results, it follows that the assumption of uncorrelated residuals is a desirable one, but not a necessary one. It remains to be seen how the residual correlation matrix derived from the residuals of the benchmark model compares to that of the restricted and unrestricted specifications.

8.9. CHAPTER SUMMARY AND CONCLUSION

This chapter introduces the benchmark model that is hypothesised to represent an adequately specified linear factor model against which restricted and unrestricted specifications may be compared in the chapters that follow. The benchmark model comprises the macroeconomic factors and two residual market factors, which are shown to imperfectly proxy for the pervasive factors in returns (Section 7.4.; Table 7.7.). A factor analytic augmentation is also incorporated to account for any omitted and unobserved factors that are not reflected by the pre-specified factors.

The results in Section 8.3. indicate that the proportion of statistically significant coefficients and explanatory power is comparable to notable studies conducted on the US stock market. Although the South African stock market differs significantly, these studies provide an imperfect benchmark for establishing what constitutes an acceptable number of significant coefficients and acceptable explanatory power (Section 8.3.1.; Section 8.3.3.). Importantly, the estimated relationships between industrial sector returns and the macroeconomic factors are in line with *a priori* expectations. This lends credence to the theoretical basis of the benchmark model. Interestingly, the factors that comprise the factor analytic augmentation, f_1 and f_2, are widely significant across sectors suggesting that there are unobserved and unspecified factors that impact returns in addition to the macroeconomic factor set and the residual market factors. This can be interpreted as preliminary evidence that models that include macroeconomic factors and follow the approach of using a residual market factor or even two residual market factors may remain underspecified (Section 8.3.2.).

Model diagnostics and robustness checks do not point towards misspecification or other econometric issues although the residuals of a number of sectors exhibit statistically significant residual serial correlation. Nevertheless, this does not appear to impact the overall results (Section 8.4.). ARCH/GARCH modelling indicates that the variance structures of most series are described by the short-memory ARCH(1) model (Section 8.5). Factor analysis of the benchmark residual correlation matrix suggests that the single extracted factor is sector-specific and transient in nature (Section 8.7.). A direct analysis of the residual correlation matrix provides further support for this. This analysis indicates that residual correlation coefficients are of a generally low magnitude and the vast majority of coefficients are statistically insignificant. In instances where correlation coefficients are statistically significant, there is no clear pattern that emerges. Co-movement does not appear to be extensively positive or negative. The Jennrich (1970) test of equality between matrices reveals that the benchmark residual matrix is not an approximation of an identity matrix. This implies that the underlying assumption of uncorrelated residuals that underlies the linear factor model may be a convenient theoretical construct but is not achievable in practice (Section 8.8.).

In summary and conclusion, the benchmark model appears to be a valid and well-specified description of the return generating process. Adequately and well-specified models should be comparable to this model across the numerous aspects considered. Whether this is the case for the restricted model and the unrestricted models is investigated in Chapter 9 and Chapter 10 respectively.

CHAPTER 9

MACROECONOMIC FACTORS AND UNDERSPECIFICATION IN THE LINEAR FACTOR MODEL

9.1. INTRODUCTION

Chapter 8 develops a specification that is assumed to represent an adequately specified linear factor model in terms of the APT framework. This specification comprises the macroeconomic factors shown in Chapter 7 to proxy for the pervasive influences in South African stock returns (Section 7.4.), the two residual market factors and a factor analytic augmentation. The results in Chapter 8 provide indirect preliminary evidence that suggests that any specification that excludes the residual market factors and the factor analytic augmentation may be underspecified.

This chapter investigates the adequacy of the linear factor model when only the macroeconomic factors are incorporated into a restricted specification. The adequacy, or the lack thereof, is established by considering the impact of the omission of the residual market factors and the factor analytic augmentation on the overall estimation results, model coefficients, model diagnostics, residual variance and the conditional variance structures and model predictions – aspects that will be impacted by factor omission, as outlined in Section 5.3.1. and also discussed in Section 5.4. It is of course possible that a factor structure that comprises macroeconomic factors adequately approximates the linear factor model. To establish this, the benchmark model is used as a comparative specification. The parameters and results of the restricted specification are compared to those of the benchmark model to establish its adequacy and the impact of the exclusion of factors. In the absence of any discernible differences that are attributable to factor omission, the restricted specification may be deemed adequate. If this is the case, then the conclusion is that macroeconomic factors are sufficient to provide an adequate and valid description of the linear factor model. However, the results of the factor regressions in Section 7.4. indicate that macroeconomic factors are poor proxies for the pervasive factors in returns.

This chapter proceeds as follows; Section 9.2. outlines the restricted specification and Section 9.3. provides a general overview of the results and discusses the impact of factor omission on the results. Section 9.4. sets out the regression diagnostics and considers the distribution of the conditional error terms. The impact of factor omission on residual variance and the conditional variance structure is considered in Section 9.5. Section 9.6.

investigates the impact of factor omission on the resultant mean errors and measures of predictive accuracy. In Section 9.7., a factor analytic investigation of the residual correlation matrix is undertaken and comparisons of the structure of the resultant residual correlation matrix take place in Section 9.8. Section 9.9 concludes the chapter.

9.2. RESTRICTED MODEL SPECIFICATION

The investigation of the ability of macroeconomic factors to proxy for the underlying pervasive factors in returns and their ability to adequately represent the linear factor model follows from the following specification:

$$\begin{aligned} R_{it} &= \alpha + b_{iBP}BP_{t-1} + b_{iLEAD}LEAD_{t-1} + b_{iBUS}BUS_t + b_{iUSD\varepsilon}USD\varepsilon_t + b_{iMET}MET_t + b_{iLTY}LTY_t \\ &+ b_{iTLI}TLI\varepsilon_t + \varepsilon_{it} \end{aligned} \tag{9.1}$$

As in equation (8.1), all the factors remain the same; R_{it} is the return on industrial sector index i at time t and the b_i's are the sensitivities to innovations in the respective macroeconomic factors, BP_{t-1}, $LEAD_{t-1}$, BUS_t, $USD\varepsilon_t$, MET_t, LTY_t and $TLI\varepsilon_t$. The residuals are represented by ε_{it}. This specification excludes the residual market factors and the factor analytic augmentation. As with the benchmark specification, this specification is estimated using ML estimation as outlined in Section 6.4.2. Initially, the conditional variance underlying each series is assumed to follow the same ARCH(*p*) or GARCH(*p,q*) process as in the benchmark model. However, the number of ARCH and/or GARCH terms is increased as required until the residuals are free of non-linear dependence and heteroscedasticity (Armitage & Brzeszczyński , 2011: 1529). The aim of the restricted specification is to establish whether a purely macroeconomic factor set can adequately describe the return generating process, without the inclusion of the residual market factors. The impact of the inclusion of $M\varepsilon_t$ and subsequently $IM\varepsilon_t$ on the results and other associated aspects is considered in Chapter 10.

9.3. MODEL OVERVIEW AND COMPARISONS

9.3.1. Macroeconomic Factor Significance Comparisons

The abridged results of the restricted specification are reported in Table 9.1.

Table 9.1: Summary Of Restricted Model Results

Panel A: Coefficient And Significance Summary							
Factor	**Mean Coeff.** **Std Error** **|Z-score|**	**Mean LS Co.** **| Diff. |**	$b_{ik} > 0$	$b_{ik} = 0$	$b_{ik} < 0$	**Total Sig.**	**ΔSig**
Intercept	0.008▲*** (0.004)▲ [1.874]▼	0.006 0.002▲***	16	10		16	+3
BP_{t-1}	0.025▼*** (0.042)▲ [0.922]▼	0.038 0.014▲***	3	23		3	-7
$LEAD_{t-1}$	0.837▼ (0.537)▲ [1.609]▼	0.883 0.046▲	12	14		12	-5
BUS_t	0.078▼ (0.049)▲ [1.667] ▼	0.079 0.0004▼	13	13		13	-5
$USD\varepsilon_t$	-0.167▲ (0.156)▲ [1.422] ▼	-0.174 0.007▲	-	16	10	10	-6
MET_t	0.156▲ (0.099)▲ [1.656] ▼	0.173 0.016▼w	12	13	1	13	-1
LTY_t	-4.037▼ (1.611)▲ [2.934]▼	-3.727 0.311▲w	-	6	20	20	0
$TLI\varepsilon_t$	3.050▲** (1.101)▲ [2.871]▼	3.076 0.026▼	22	4	-	22	-2

Panel B: Goodness-of-fit and Model Selection Criteria			
	Mean Value	**Minimum**	**Maximum**
$\bar{R}^2$	0.142▼***	0.032 Fixed line telecom.	0.255 Life insurance
AIC	-2.736▲***	-3.597 Food producers	-1.434 Ind. metals & mining sector
BIC	-2.556▲***	-3.411 Food producers	-1.248 Industrial metals & mining

Table 9.1: Summary Of Restricted Model Results (Continued…)

Notes:
The asterisks, ***, ** and *, indicate statistical significance at the respective 1%, 5% and 10% levels of significance. All factors are in innovations (unexpected changes) (Section 6.2.2; Section 6.2.3.; Table 6.4.), where BP_{t-1} - Building Plans Passed, $LEAD_{t-1}$ - Leading Indicator, BUS_t - Business Activity, USD_t - Rand-Dollar Ex. Rate, MET_t - Metal Prices, LTY_t - Long-Term Gov. Bond Yields, TLI_t - Trading Partner Lead. Index, R_{Mt} - JSE All Share Index and R_{IMt} - MSCI World Index (US$). In Panel A, Mean Coeff. is the mean value of the intercept and the coefficients associated with each factor. Values in the parentheses ( ) are the mean coefficient standard errors (Std Error) and the values in the brackets [ ] are the mean *z*-scores (|Z-score|). In the third column, Mean LS Co. are the mean values of least squares intercepts and coefficients of the benchmark model. |Diff.| are the absolute mean differences between the ML and least squares coefficients. $b_{ik} > 0$ and $b_{ik} < 0$ indicate the respective number of coefficients that are statistically significant and have a positive or negative impact. Total Sig. is the total number of statistically significant coefficients associated with each factor across the return series in the sample. **Δ**Sig is the increase or decrease in the number of statistically significant coefficients relative to the benchmark model specification. In Panel B, Mean is the arithmetic mean of the $\bar{R}^2$, AIC and BIC values across sectors. The Minimum and Maximum values correspond to the lowest and highest values observed and the associated sectors for which they are observed. Across panels, the first ▲ or ▼ symbol indicates that a value is larger or smaller relative to that observed for the benchmark model. Accompanying asterisks, if present, indicate that differences are statistically significant. Throughout, the superscript "W" indicates that the Wilcoxon matched-pairs signed-rank test contradicts the results of the paired-sample *t*-test.

The investigation into the impact of factor omission in the restricted model begins with a comparison of the number of statistically significant coefficients for the seven macroeconomic factors in the restricted model to that of the benchmark. While 119 of the 182 (65.38%) estimated macroeconomic factor coefficients are statistically significant for the benchmark model, 93 of the 182 (51.099%) estimated macroeconomic factor coefficients are statistically significant for the restricted specification. This represents a decrease of 26 (a 14% decrease) in the number of statistically significant coefficients for the macroeconomic factors. Such an outcome is predicted by the literature and can be attributed to inflated standard errors (Section 5.3.1.; Section 6.4.6.; discussed in Section 9.5.). Sykes (1993: 26) shows that the omission of relevant factors is associated with increased standard errors and lower *t*-statistics. This is the reason for Van Rensburg's (2002: 91) use of two residual market factors and it is acknowledged that underspecification will translate into an upward bias in coefficient variance leading to erroneous instances of not rejecting the null hypothesis of an absence of a statistically significant relationship. The comparison of the number of statistically significant coefficients for the macroeconomic factors in the restricted model and the benchmark model suggests that underspecification leads to an erroneous understatement of the importance of factors in the return generating process.

Intercepts are now also statistically significant for three additional industrial sectors. A consequence of underspecification is that the intercepts will now be biased and will reflect the mean effect of the omitted factors (Sykes, 1993: 25). Such a result is therefore expected if factors are omitted. Factors that now appear to be less important are BP_{t-1} (decreasing from 10 statistically significant instances to three statistically significant instances [-7]), $LEAD_{t-1}$ (17 to 12 [-5]), BUS_t (18 to 13 [-5]), $USD\varepsilon_t$ (16 to 10 [-6]), MET_t (14 to 13 [-1]) and $TLI\varepsilon_t$ (24 to 22 [-2]). While BUS_t, MET_t and $TLI\varepsilon_t$ may still be considered as systematic in nature, their importance is understated. This is especially true for BUS_t which is now statistically significant for 13 sectors in the restricted specification as opposed to 18 sectors in the benchmark model. The factors that are most impacted by the decreases in the number of significant instances are BP_{t-1}, $LEAD_{t-1}$ and $USD\varepsilon_t$. While BP_{t-1} is found to be statistically significant for 10 industrial sectors in the benchmark specification, suggesting that although it may not be a truly systematic factor (contrary to the factor-return correlation analysis in Table 7.4. and the factor score regressions in Table 7.7.) and therefore is not a true candidate APT factor, its impact in the restricted version of the linear factor model is even more severely understated. Szczygielski and Chipeta (2015: 15) suggest that this factor is systematic in nature. However, the results in Table 9.1. now indicate that the impact of BP_{t-1} is limited to only three sectors (the food and drug retailers, general retailers and equity investments and instruments industrial sectors). The results for $LEAD_{t-1}$ and $USD\varepsilon_t$ are also noteworthy. The APT framework requires that systematic factors with a pervasive impact feature in the linear factor model (Burmeister *et al.,* 1994: 3). These two factors no longer appear to be pervasive. The impact of both factors is now confined to fewer than half of the industrial sectors. Ferson and Harvey (1994: 785) state that factors with insignificant betas should be excluded from further analysis and therefore, cannot be associated with significant expected return premia in the APT relation. The consequence is the erroneous exclusion of these seemingly non-systematic factors and a misspecification of the linear factor model and the APT relation (Elton *et al.,* 1995: 1239; Section 5.4.2.).

The results of this section suggest that underspecification results in an understatement of the importance of certain macroeconomic factors in the linear factor model and therefore a misidentification of the return generating process. This understatement will carry over into the APT relation.

9.3.2. Coefficient Magnitude Comparisons

The discussion now turns to the magnitudes of the estimated model parameters; the intercept and coefficients which can be compared to those of the benchmark model in Table 8.1. The results in Table 9.1. indicate that the mean intercept value increases from 0.006 for the benchmark model to 0.008 for the restricted model (as indicated by ▲). This is expected if factor omission introduces bias into the intercept. The paired sample *t*-test confirms that the means of the intercepts obtained from these two models differ significantly. This finding is similar to that of Lehmann and Modest (1987: 259), who suggest that intercepts are sensitive to the factor structure of the linear factor model. While Lehmann and Modest (1987) report somewhat ambiguous results, showing that there is a differing impact on the magnitude of the mean intercepts, depending upon subperiods considered, the results presented here indicate that the intercepts are significantly impacted, even if the impact appears to be small in magnitude. These results suggest that in the presence of underspecification, any inferences drawn from the magnitude of the alphas of the linear factor model relating to performance may be misleading (see also Van Rensburg, 2002: 92).

The paired-sample *t*-test is also applied to determine whether the mean coefficients on the macroeconomic factors differ across the two specifications. Results indicate significant differences between the coefficient series for BP_{t-1} and $TLI\varepsilon_t$. The mean of the coefficients for BP_{t-1} decreases from 0.037 in Table 8.1. to 0.025 in Table 9.1. and increases for $TLI\varepsilon_t$ from 2.865 to 3.050. This suggests that the exclusion of factors has a significant impact on the coefficient estimates for these two factors. This is also to be expected; coefficients either underestimate or overestimate the true impact of a factor if a model is underspecified. This is true if the factor set is correlated with the omitted factors, which is not the case if the omitted factors, $M\varepsilon_t$, $IM\varepsilon_t$, f_{1t} and f_{2t}, are orthogonal by construction (Wooldridge, 2013: 90). However, as this study relies upon the ARCH/GARCH framework to estimate the underlying conditional variance, it is likely that the omission of these factors impacts the structure of heteroscedasticity. Consequently, given the econometric methodology that is applied, the change in the structure of heteroscedasticity is reflected in the coefficients (Hamilton, 2010; Armitage & Brzeszczyński, 2011: 1526). All the coefficient means for the remaining factors are also different in magnitude from those of the benchmark model. However, neither the paired-sample *t*-test nor the Wilcoxon matched-pairs signed-rank test indicate that differences are statistically significant.

Reported in Table 9.1., in the third column, are the absolute mean differences between the ML coefficients of the restricted specification and those of the benchmark model estimated using least squares. As suggested by Sweeney and Warga (1986: 38) and Jorion (1991: 366), biases in the coefficient estimates (the betas) that feature in the APT relation as data (Section 2.2.) have the potential to result in erroneous inferences relating to the nature of priced factors. The results indicate that mean differences between the ML and least squares coefficient estimates for the restricted specification are generally greater than those for the benchmark specification (as indicated by ▲). For example, the absolute mean differences between the ML and least squares coefficients are now statistically significant for BP_{t-1} (t-test), MET_t and LTY_t (Wilcoxon) although the differences for estimates of coefficients for $USD\varepsilon_t$ (Wilcoxon) and $TLI\varepsilon_t$, are no longer statistically significant. What is perhaps more telling is the actual magnitude of the differences. In certain instances, the deviation from least squares coefficients is notable. For example, the absolute mean difference for the intercept increases from 0.0006 in Table 8.1. to 0.002 in Table 9.1. Moreover, the absolute mean differences more than double for BP_{t-1} (0.001 to 0.014), and $LEAD_{t-1}$ (0.022 to 0.046), and increase for all remaining factors with the exception of BUS_t, MET_t and $TLI\varepsilon_t$ for which the differences decrease (as indicated by ▼). It appears that the deviation from least squares coefficients increases for four of the seven macroeconomic factors although differences between the least squares coefficients of the benchmark model and the ML coefficients of the restricted model are not always significant. Nevertheless, these results show that when factor omission is quantified by taking into account the structure of heteroscedasticity, coefficient estimates will reflect bias and will be inconsistent. A potentially analogous result will be obtained if the omitted factors (the two residual market factors and the factors comprising the factor analytic augmentation) are correlated with the macroeconomic factors in the restricted specification.

In summary, factor omission impacts parameter estimates, not least the intercepts which increase in magnitude, but also the coefficient estimates associated with the macroeconomic factors. The differences between ML and least squares coefficients generally increase suggesting that coefficients will reflect greater levels of bias (consequence 1) in Section 5.3.1.). This is potentially attributable to the introduction of impure heteroscedasticity in the conditional variance, as reflected by the ARCH(p) and GARCH(p,q) specifications used to model conditional variance (discussed further in Section 9.5.). Poon and Taylor (1991: 318) aptly summarise the impact of coefficient bias by stating that tests of the APT relation are in terms of the true value of risk, the coefficients, yet empirical tests rely upon coefficient estimates, which will be biased. These biases compound the error-in-variables problem, given that inputs for the APT relation are derived from the linear factor model.

9.3.3. Model Assessment And Comparisons

Also impacted are the respective $\bar{R}^2$ s and the AIC and BIC values. The mean $\bar{R}^2$ decreases from 0.504 for the benchmark specification to 0.142 for the restricted specification indicating that the omission of the residual market factors and statistically derived factors results in a significant decrease in the explanatory power of the specification. The paired sample *t*-test confirms the statistical significance of the difference between the mean $\bar{R}^2$ s of the benchmark and restricted specifications (this is also confirmed by the Wilcoxon matched-pairs signed-rank test). This low mean $\bar{R}^2$ for the macroeconomic factor set is comparable to the $\bar{R}^2$ of 0.109 in Connor (1995: 44) for a sample of 799 high capitalisation US stocks. Although, Connor (1995) relies upon a smaller set of macroeconomic factors that are specific to the US market, namely inflation, term structure, industrial production, unemployment and the junk bond premium, the results are comparable. Both sets of results suggest that macroeconomic factors, by themselves, are poor proxies for pervasive influences in stock returns.

A comparison of the range of the $\bar{R}^2$ s confirms that all sectors are impacted; the $\bar{R}^2$ ranges between 0.255 for the life insurance sector to 0.032 for the fixed line telecommunications sector (Panel B of Table 9.1.). For the benchmark specification, the $\bar{R}^2$ ranges between 0.941 for the mining sector and 0.171 for the fixed line telecommunications sector (Panel B of Table 8.1.). The statistically significant decrease in the overall $\bar{R}^2$ not only demonstrates the impact of underspecification on the ability of a model to explain the proportion of the total variation in returns, it also highlights the drawbacks of using a single specification to describe multiple series. A single specification may be unsuitable for a broader sample of assets (Bilson *et al.,* 2001). Perhaps the aspect that is most concerning is that while the benchmark specification explains, on average, over half of the total variation in returns ($\bar{R}^2$ >0.5), the restricted specification explains, on average, just over a 10th of the variation in returns. The contribution of the macroeconomic factors to explaining return behaviour appears almost negligible.

The mean AIC statistic is -2.736 for the restricted specification. The minimum value is -3.597 for the food producers sector and the maximum value is -1.434 for the industrial metals and mining sector. This is a marked contrast to the mean AIC value and the range of AIC values for the benchmark model. The respective AIC values are lower for the benchmark model (mean AIC statistic is -3.348 and the respective extreme AIC statistics values are -4.956

and -2.035 for the mining and fixed line telecommunications sectors) relative to those obtained from the restricted specification (Panel B of Table 8.1.). This indicates that the ability of the restricted model to replicate the actual return observations is inferior to that of the benchmark specification (Spiegelhalter *et al.,* 2014: 1-2). Similarly, the mean BIC statistic of -2.556 for the restricted specification is significantly higher than that for the benchmark specification of -3.114. As the BIC statistic attempts to identify the specification that most closely resembles the true return generating process, these results indicate that the restricted specification is less likely to approximate the true return generating process in comparison to the benchmark model (Aho, Derryberry & Peterson, 2014: 635). A comparison of the minimum and maximum BIC values leads to a similar conclusion; both are higher relative to those of the benchmark specification. The highest BIC value is -1.248 for the industrial metals and mining sector (compared to -1.797 for the fixed line telecommunications sector in Panel B of Table 8.1.) and the lowest BIC value is -3.411 for the food producers sector (compared to -4.736 for the mining sector in Panel B of Table 8.1.).

In conclusion, explanatory power, as measured by the $\bar{R}^2$ decreases significantly with the omission of relevant factors. The higher AIC statistics indicate that predictive accuracy significantly deteriorates whereas the higher BIC statistics indicate that the restricted model significantly deviates from an appropriate representation of the true return generating process.

9.4. MODEL DIAGNOSTICS AND ROBUSTNESS

Table 9.2. reports the abridged results of the model diagnostic tests for the restricted specification. The fourth column reports the number of changes in significant observations relative to the benchmark specification.

Table 9.2: Abridged Restricted Model Diagnostics

Test	Mean Value	Total Sig.	ΔSig
F-Test	6.692▼	26/26	-
JB Test	15.868▲	14/26▼	-1
$Q(1)$	1.919▲	8/26▲	+3
$Q(5)$	7.107▲	8/26▲	+4
$Q^2(1)$	0.524▲	0/26	-
$Q^2(5)$	3.638▼	0/26	-
ARCH(1)	0.454▲	0/26	-
ARCH(5)	0.724▼	0/26	-

Table 9.2: Abridged Restricted Model Diagnostics (Continued…)

Notes:
Significance is recorded at the 10% level of significance. *F*-Test reports the results for Wald's test of linear restrictions jointly equating all explanatory factors in the respective specifications to zero. JB Test summarises the results of the Jarque-Bera test for normality. $Q(1)$ and $Q(5)$ are Ljung-Box *Q*-statistics indicating whether serial correlation in the residuals is statistically significant at the first order and jointly up to five orders of serial correlation respectively. $Q^2(1)$ and $Q^2(5)$ are Ljung-Box test statistics for non-linear dependence in the residuals at the first order and jointly up to five orders. ARCH(1) and ARCH(5) are Lagrange Multiplier (LM) tests for ARCH effects in the residuals at the first and fifth orders respectively. Mean Value reports the mean of the respective test statistics and Total Sig. reports the number of instances in which the results of the respective tests applied are statistically significant. **Δ**Sig is the increase or decrease in the number of statistically significant coefficients relative to the benchmark model specification. The first ▲ or ▼ symbol indicates that a value is larger or smaller relative to that observed for the benchmark model.

As with the benchmark specification, the null hypothesis for Wald's test restricting all coefficients in equation (9.1) to zero is rejected for all industrial sectors. This confirms the overall significance of the macroeconomic factors by themselves in the restricted model. This outcome is expected; the results in Panel A of Table 7.7. confirm that the macroeconomic factors are proxies for factors impacting stock returns. Notably, the mean *F*-statistic decreases from 36.528 for the benchmark model in Table 8.2. to 6.692 for the restricted model in Table 9.2. This implies that the sum of squared residuals for the restricted model is greater than that of the benchmark model (SSR_{ur} in equation (6.30) where SSR_{ur} is the sum of squared residuals for the benchmark model in the present context). This will be the case if a lower proportion of return variation is explained by the restricted model in equation (9.1). The effect of this will be to lower the value of *F*-statistics across sectors, as reflected in Table 9.2. (Blackwell, 2008: 4). Indirectly, the *F*-statistics reflect a decrease in the explanatory power of the restricted specification (Kluve, Schneider, Uhlendorff & Zhao, 2012: 600).

The exclusion of factors does not have much of an impact on conditional normality. While the residuals of 15 sectors in Table 8.2. are associated with significant JB test statistics, 14 residual series are associated with significant JB test statistics in Table 9.2. As the number of instances of significant departures from normality is comparable, these results are somewhat ambiguous. An analysis of individual residual series is also not revealing (see Table A1.2. in Appendix A). While some series are no longer non-normally distributed, for example the mining, construction and general industrials residual series, others, such as the electronic and electrical equipment, industrial transport and general retailers sectors now exhibit significant departures from normality. Although it is difficult to pronounce whether the results for individual sectors are attributable to the exclusion of the residual market factors and the factor analytic augmentation from the restricted model, the mean JB statistic suggests that non-normality has increased across sectors. The mean JB statistic, which is jointly determined by the level of skewness and kurtosis (equation (6.4)), is 15.868 for the restricted model but 11.511 for the benchmark specification. This suggests that overall, the levels of kurtosis and/or skewness have increased, even if not significantly (Varga & Rappai, 2002: 134). As proposed by Downing and Clark (2010: 403), this is potentially attributable to the presence of omitted factors that would, had they been included, explain outliers which contribute to departures from normality.

The $Q(1)$ and $Q(5)$ statistics in Table 9.2. present a somewhat clearer picture. The mean $Q(1)$ and $Q(5)$ statistics for the restricted model are 1.919 and 7.107 respectively. These values are marginally higher than the respective Q-statistics of 1.744 and 6.586 for the benchmark model in Table 8.2. This suggests that overall, the level of joint serial correlation in the residuals of the restricted specification is higher relative to that for the benchmark model (equation (6.6)). Q-statistics for the benchmark model indicate that the residuals of a total of six sectors exhibit statistically significant serial correlation, as evident from either or both statistically significant $Q(1)$ and $Q(5)$ statistics (see Table A1.1. in Appendix A). In contrast, the residuals of the restricted specification are significantly serially correlated for almost half of the industrial sectors; a total of 12 industrial sectors exhibit serially correlated residuals at either lower order serial correlation (first order serial correlation) or higher order serial correlation (up to the fifth order) or both. Industrial sectors for which the residuals are also now serially correlated at either of the orders are the forestry and paper, industrial metals and mining, mining, construction and materials, general industrials, pharmaceuticals and biotechnology, media and banks industrial sectors. These results indicate that the restricted specification is more likely to exhibit residual serial correlation - potentially impure serial correlation induced by underspecification (Mutsune, 2008: 6). To investigate whether serial correlation is sufficiently large so as to distort the results for the sectors that now exhibit serial correlation, the restricted specification is re-estimated using the least squares methodology with HAC standard errors. Results indicate that there are minor changes in factor significance for some sectors. For the metals and mining industry, $TLI\varepsilon_t$ is no longer statistically significant. For the construction and materials and the banks industrial sectors, $TLI\varepsilon_t$ is now statistically significant. For the general industrials sector, BUS_t is now statistically significant. This suggests that the induced serial correlation may have some, albeit minor, impact on inferences.

As with the benchmark specification, the $Q^2(1)$ and $Q^2(5)$ statistics indicate that the residuals do not exhibit non-linear serial correlation. Tests for ARCH(1) and ARCH(5) effects in the residuals confirm that residuals are free of ARCH effects indicating that residual variance is stationary. As the approach employed relies upon increasingly complex ARCH(p) and GARCH(p,q) specifications to ensure that residual series are free of non-linear dependence and heteroscedasticity (Section 6.4.2.), it is expected that residuals will be free of non-linear dependence and heteroscedasticity. However, this comes at the cost of the need to apply the more complex GARCH(p,q) specification to a greater number of series, which will become evident in Section 9.5.

The observations in this section suggest that the impact of factor omission on the conditional distribution of the residuals is ambiguous. Deviations from normality may be statistical artefacts that are unrelated to underspecification. In terms of overall model significance, factor omission leads to lower F-statistics, although macroeconomic factors continue to provide a valid multifactor description of the return generating process. Factor omission appears to introduce impure residual serial correlation. This seemingly induced serial correlation has a minor impact on inferences relating to the statistical significance of factors.

9.5. VARIANCE AND CONDITIONAL VARIANCE

The results in Table 9.1. provide preliminary evidence that factor omission translates into a loss of parameter efficiency, which is attributable to an upward bias in residual variance (Section 6.4.6.). An upward bias in residual variance translates into erroneous failures to not reject the null hypothesis of no impact as a result of biased standard errors, lower critical values and wider confidence intervals (Sykes, 1993: 16; Van Rensburg, 2000: 37; 2002; 91; Studenmund, 2014: 178-200; consequence 4) in Section 5.3.1.). The consequences of inflated residual variance stemming from underspecification are therefore two-fold. The linear factor model may be misidentified, as suggested by the comparison of the restricted and benchmark models in Section 9.3.1. Also, if residual variance or standard deviation is used in tests of the validity of the APT relation, the APT may erroneously be declared invalid.

The results in Panel A of Table 9.1. point towards inflated standard errors. The mean standard errors for the intercepts and coefficients on the macroeconomic factors for the restricted model are higher than those of the benchmark model in Panel A in Table 8.1. For example, the mean standard error for BP_{t-1} increases from 0.031 in Panel A of Table 8.1. to 0.042 in Panel A of Table 9.1., an increase of 35.484%. This is also true for the remaining factors and it appears that the increases in standard errors are more than proportional to increases in coefficients even for factors for which coefficients have increased in magnitude, namely $USD\varepsilon_t$, MET_t and $TLI\varepsilon_t$. That the standard errors have increased in general and increased more than proportionally to the coefficients for these factors is evident from lower (absolute) mean *z*-scores. For example, the mean z-score for $USD\varepsilon_t$ decreases from 2.075 to 1.422 (-31.470%), for MET_t, the mean z-score decreases from 2.410 to 1.656 (-31.286%) and for $TLI\varepsilon_t$, the mean z-score decreases from 3.809 to 2.871 (-24.626%). Although the mean z-score decreases from 2.013 to 1.874 (-6.905%) for the intercept, this is the only parameter that is associated with an increase in the number of significant instances. This suggests the increase in the standard errors is of an insufficient magnitude to translate into a failure to reject the null hypothesis for the intercepts. For the intercepts, the mean standard errors increase from 0.003 in Panel A of Table 8.1. to 0.004 in Panel A of Table 9.1. Percentage decreases in the mean absolute *z*-scores for the macroeconomic factors range between 37.449% for BP_{t-1} to 23.881% for BUS_t and it is evident that these decreases are partially driven by increases in coefficient standard errors (also see Kluve *et al.,* 2012: 600). These observations point towards increases in residual variance, which constitutes preliminary evidence of a loss of parameter efficiency arising from factor omission. The following discussion shows that residual variance increases overall, providing support for this argument.

The mean residual variance, $\sigma^2_{RES\varepsilon_i}$, for the restricted model reported in Panel A of Table 9.3., is significantly higher than that of the benchmark specification, $\sigma^2_{BM\varepsilon_i}$, as confirmed by a paired-sample *t*-test. The mean residual variance for the restricted specification is 0.004307. This is almost double the mean residual variance of 0.002483 in Panel A of Table 8.3. for the benchmark specification. The sector with the lowest residual variance is the food producers sector (0.001469) and the sector with the highest residual variance (0.014494) is the industrial metals and mining sector. The sector with the lowest level of residual variance for the benchmark specification is the mining sector (0.000368) whereas the sector with the highest level of residual variance is the software and computer services sector (0.00682). This points towards an upward shift of the lower and upper values of residual variance for the restricted model (see Table A1.2. in Appendix A).

Table 9.3: Restricted Model Residual Variance And Variance Structure

	Mean Value	Minimum	Maximum
Panel A: Residual Variance			
	Mean Value	**Minimum**	**Maximum**
$\sigma^2_{RES\varepsilon_i}$	0.004307▲***	0.001469 Food producers	0.014494 Industrial metals & mining
Sig. $\sigma^2_{RES\varepsilon_i} > \sigma^2_{BM\varepsilon_i}$	22/26		
Sig. $\sigma^2_{RES\varepsilon_i} = \sigma^2_{BM\varepsilon_i}$	4/26		
Sig. $\sigma^2_{RES\varepsilon_i} < \sigma^2_{BM\varepsilon_i}$	0/26		
Panel B: Conditional Variance Structure			
Model	**ARCH(1)**	**ARCH(2)**	**GARCH(1,1)**
Frequency	10	1	15
	Mean Coeff.	**Mean Coeff.**	**Mean Coeff.**
ω	0.004 (10)	0.001 (1)	0.0005 (3)
α_1	0.120 (2)	0.082	0.095 (5)
α_2		0.134	
β_1			0.776 (15)
Sig. *F*-Test	**2/10**	**0/1**	**15/15**
Panel C: Conditional Heteroscedasticity			
α_i	0.109▲		

Notes:
In Panel A, the Mean Value is the mean of the residual variance across sectors. The Minimum and Maximum values are the lowest and highest residual variances observed for the respective sectors. Sig. $\sigma^2_{RES\varepsilon_i} > \sigma^2_{BM\varepsilon_i}$, $\sigma^2_{RES\varepsilon_i} = \sigma^2_{BM\varepsilon_i}$ and $\sigma^2_{RES\varepsilon_i} < \sigma^2_{BM\varepsilon_i}$ are comparisons of the residual variance derived from the restricted model and the benchmark model. The number of significant instances as established by the Brown-Forsythe test is reported as is the number of instances in which no significant differences are observed. In Panel B, Frequency is indicative of the number instances of each ARCH(*p*) or GARCH(*p,q*) model applied. Negative ARCH coefficients, which are indicative of the absence of conditional heteroscedasticity and are therefore statistically insignificant, are rounded to zero in aggregation. This is consistent with the non-negativity constraint ($\alpha_i \geq 0$) placed upon the ARCH coefficient in the ARCH(*p*) specification (Poon, 2005: 38). The numbers in brackets ( ) next to each mean value indicate the number of statistically significant coefficients for each ARCH(*p*) or GARCH(*p,q*) specification at the 10% level of significance. Sig. *F*-Test reports the number of significant instances of Wald's test of linear restrictions for the ARCH and GARCH coefficients. The null hypothesis is that ARCH and GARCH coefficients are jointly equal to zero. In Panel C, α_i is the arithmetic mean of the ARCH coefficients across sectors. Across panels, the first ▲ or ▼ symbol indicates that a value is larger or smaller relative to that observed for the benchmark model. Accompanying asterisks, if present, indicate that differences are statistically significant. A superscript "W" indicates a discrepancy between the results of the *t*-test and the Wilcoxon test.

The Brown-Forsythe test is applied to determine whether the variance of the residuals for each sector is homogeneous across the benchmark and restricted specifications (Section 6.4.6.). The null hypothesis of homogeneous variance is rejected for all industrial sectors with the exception of the automobiles and parts, the pharmaceuticals and biotechnology, fixed line telecommunications and the software and computer services industrial sectors ($\sigma^2_{RES\varepsilon_i} = \sigma^2_{BM\varepsilon_i}$). Results indicate that the residual variance of the restricted model for the individual sectors is significantly higher than that of the residuals of the benchmark specification for 22 out of 26 industrial sectors ($\sigma^2_{RES\varepsilon_i} > \sigma^2_{BM\varepsilon_i}$). Notably, the abovementioned industrial sectors are also associated with relatively low $\bar{R}^2$s in the benchmark model. This explains why the exclusion of the residual market factors and the factor analytic augmentation has little impact on residual variance for these sectors. These factors contribute little to explaining the variation in returns for these sectors in the first instance. In contrast, the mining sector has the highest $\bar{R}^2$ (0.941) in the benchmark model and this decreases to 0.167 in the restricted model. With the exclusion of factors, the residual variance for this sector increases from 0.000368 in the benchmark specification to 0.005324 in the restricted specification. This represents an almost 14.5-fold increase in residual variance for this sector. Another example is the general retailers sector which has an $\bar{R}^2$ of 0.779 for the benchmark model and an $\bar{R}^2$ of 0.246 for the restricted model. For this sector, residual variance increases 3.5-fold, from 0.000907 to 0.003169. Although the increase in variance is not as dramatic as for the mining sector, it supports the hypothesis that the exclusion of factors, which contribute significantly to explaining return variation significantly, inflates residual variance (Lehmann, 1990: 72; Dominguez, 1992: 97, 98).

This analysis of the magnitude of residual variance in Panel A of Table 9.3. also provides further direct support for the discussion in Section 9.3.1., which attributes the understated impact of a number of macroeconomic factors in Panel A of Table 9.1. to the loss of efficiency. The larger standard errors reported in Table 9.1. relative to those in Table 8.1., together with the results reported in Panel A of Table 9.3., point towards unreliability in drawing inferences relating to the significance of certain factors. Factor omission is associated with a loss of efficiency, erroneous inferences and may lead to erroneous rejections of the APT. The increases in mean standard errors suggest that this is indeed the case; there is a general loss of efficiency following the omission of the residual market factors and the factor analytic augmentation.

The results in Panel B of Table 9.3. indicate that the structure of the conditional variance underlying the restricted model is more complex relative to that of the benchmark specification. The conditional variance for 10 sectors is described by the short-memory ARCH(1) process with the *F*-statistic statistically significant for two sectors. These results indicate that the ARCH(1) model captures time-varying volatility for two industrial sectors, namely the electronic and electrical equipment sector and the travel and leisure industrial sector (see Table A1.2. in Appendix A). This is in contrast to 18 sectors for the benchmark model, with three sectors exhibiting statistically significant *F*-statistics. An ARCH(2) model describes the conditional variance of returns on the food producers sector although the associated *F*-test indicates that both ARCH coefficients are not jointly significant. This indicates that a somewhat more complex ARCH/GARCH-type specification is required to describe conditional variance for this sector although ARCH effects are weak. Importantly, the more complex GARCH(1,1) model is used to describe the conditional variance for returns on 15 industrial sectors. This is in contrast to eight industrial sectors for which the conditional variance is described by the GARCH(1,1) model for the benchmark model. Furthermore, the *F*-test confirms that the GARCH(1,1) model is appropriate for eight sectors when returns are described by the benchmark specification, but for 15 industrial sectors when returns are described by the restricted specification. This points towards evidence of changing volatility dynamics and an increase in the complexity of the conditional variance structures of the residuals.

While the ARCH(*p*) specification is considered to be a short-memory model, the GARCH(*p,q)* model is considered to be a long-memory model that incorporates an adaptive learning mechanism, in the form of lagged conditional variance terms (h_{it-q} in equation (6.25)), and is therefore capable of capturing more complex volatility dynamics (Bollerslev, 1986: 309; Elyasiani & Mansur, 1998: 541). This seeming increase in the complexity of the conditional variance structures may be driven by impure heteroscedasticity that is associated with factor omission in the restricted model (consequence 6) in Section 5.3.1.). In this context, Webster (2013: 230) argues that heteroscedasticity is the result of omitted factors that are now reflected in the residuals. Therefore, it follows that an increase in heteroscedasticity that is attributable to omitted factors translates into a need to use more complex ARCH/GARCH-type specifications to ensure that residuals are free from non-linear dependence and ARCH effects (Section 6.4.2.).

Overall, *F*-statistics are statistically significant for 17 out of 26 sectors for both ARCH(*p*) and GARCH(*p,q*) specifications for the restricted model. In contrast, *F*-statistics are statistically significant for a total of 11 out of 26 sectors for the benchmark model. This indicates that a greater number of sectors now require ARCH and GARCH modelling to capture time-varying residual variance dynamics relative to the benchmark model. The inducement of non-constant variance, which is reflected by the total greater number of statistically significant ARCH(*p*) and GARCH(*p,q*) specifications can translate into a misidentification of the linear factor model. Greene (2012: 312)[144] argues that, in the presence of impure heteroscedasticity, conventional standard errors will substantially depart from appropriate values suggesting that any inferences relating to the significance of the coefficients will be misleading. If conventional econometric techniques are applied that do not account for non-constant residual variance, factors that are relevant for describing the return generating process may appear to be insignificant and may be potentially excluded. This will translate into an underspecified linear factor model and APT relation (Ferson & Harvey, 1994: 785).

The mean of conditional heteroscedasticity associated with the restricted model, as represented by α_i in Panel C of Table 9.3., is 0.109. [145] This is higher than that of 0.103 for the benchmark model reported in Panel C of Table 8.3. However, a paired-sample *t*-test and the Wilcoxon matched-pairs signed-rank test do not indicate that differences are statistically significant suggesting there are no differences in the magnitude of the estimated ARCH coefficients for the ARCH(*p*) and GARCH(*p,q*) specifications. Nevertheless, it is clear that factor omission impacts the structure of the conditional variance. This is confirmed by the preceding discussion which points towards the need to use the more complex GARCH(1,1) specification more frequently to describe conditional variance structures. As proposed by Bera *et al.* (1988: 212), the changing structure of the conditional variance, whether in the form of higher levels of conditional heteroscedasticity or more complex conditional variance structures attributable to factor omission, will be reflected in the coefficients on macroeconomic factors for the restricted model. This explains the general increases in the magnitude of the deviations from least squares coefficients of the benchmark model reported in the third column of Table 9.1. and also the differences in the magnitude of mean coefficients, especially those for BP_{t-1} and $TLI\varepsilon_t$ which are statistically significant in Panel A of Table 9.1.

This section indicates that factor omission is associated with larger coefficient standard errors and lower *z*-scores. This is attributable to a general upward bias in the residual variance and is confirmed by the Brown-Forsythe test across individual industrial sectors.

[144] Greene (2012: 312) states that if heteroscedasticity is uncorrelated with factors in models for large samples, least squares estimates will not be optimal but will not be misleading. However, if this assumption is incorrect, in that heteroscedasticity is related to factors in a given model, then conventional standard errors will depart from appropriate values.

[145] The mean of conditional heteroscedasticity, α_i, is the arithmetic mean of all ARCH coefficients in the conditional variance model for each sector. A single ARCH coefficient is derived for the food producers sector which is described by an ARCH(2) model by summing the respective α_1 and α_2 coefficients to obtain a single ARCH coefficient for use in the calculation of the mean α_i.

An upward bias in residual variance has the potential to translate into misleading hypothesis tests. This can explain the misidentification of relevant factors that characterise the linear factor model and their understated impact in Section 9.3.1. Moreover, tests of the APT, if the residual variance or standard deviation is used, can result in erroneous rejections of the pricing relation. Factor omission also introduces impure heteroscedasticity into the residual variance. This is suggested by increasingly complex variance structures, as indicated by the greater number of GARCH(*p,q*) specifications used. If a technique that directly takes into consideration heteroscedasticity and the structure of conditional variance in estimating model parameters is applied, coefficients are likely to reflect omitted factors.

9.6. PREDICTIVE ABILITY

The results in Panel A of Table 9.4. indicate that the mean residuals, ε_{it}, of the restricted model differ significantly from those of the benchmark model and from zero. The mean errors derived from the restricted model, in absolute terms, are 0.0019261. The mean errors from the benchmark model are 0.0005692. This represents a more than three-fold increase in the magnitude of the residuals, following the omission of the residual market factors and the factor analytic augmentation. Following Chang (1991: 387), this suggests that the restricted model is more likely to suffer from a reduction in the power of statistical tests and inferior predictive ability. This is supported by the results in Section 9.5. relating to the observed upward bias in residual variance, which goes some way in explaining the discrepancies between the number of statistically significant coefficients observed for the benchmark model and the restricted specification and the understatement of factor significance in the restricted model reported in Table 9.1.

A consequence of underspecification is that predictions are unreliable and the higher mean errors suggest that this will be the case for the restricted model. To evaluate the ability of the restricted model to replicate realised returns and the reliability of predictions, the means of the Theil *U* statistics and its components, namely the bias, the variance and the covariance proportions, are compared to those obtained from the benchmark model (consequence 5) in Section 5.3.1.).

Table 9.4: Summary Of Mean Errors And Theil's *U* Statistic For The Restricted Model

	Panel A: Mean Errors		
	Mean value		
ε_{it}	-0.0019261***▼***		
	Panel B: Theil's *U* Statistic And Decomposition		
	Mean Value	**Minimum**	**Maximum**
Theil *U*	0.633▲***	0.547 Food producers	0.767 Fixed line telecoms
Bias (U_{BIAS})	0.002355▲**	0.000001 Pharmaceutical & biotechnology	0.020372 software and computer services
Variance (U_{VAR})	0.412575▲***	0.287454 industrial transport sector	0.613429 Software and computer services
Covariance (U_{COV})	0.585071▼***	0.366200 Software and computer services	0.711488 Industrial transportation

Notes:
The asterisks, ***, ** and *, indicate statistical significance at the respective 1%, 5% and 10% levels of significance. In Panel A, the Mean Value is the respective arithmetic mean of the residual terms. A paired-sample *t*-test is applied to test the null hypothesis that the mean value of the residuals differs significantly from zero. In Panel B, the Mean Value is the arithmetic mean of respective the measures of predictive accuracy. The Minimum and Maximum are the minimum and maximum values associated with the respective measures of accuracy for the respective sectors. Across panels, the first ▲ or ▼ symbol indicates that a value is larger or smaller relative to that observed for the benchmark model. Accompanying asterisks, if present, indicate that differences are statistically significant.

The results in Panel B of Table 9.4. indicate that the mean Theil *U* statistic has increased significantly, from 0.395 for the benchmark specification to 0.633 for the restricted specification. This corresponds to a 60.253% increase in this statistic and suggests that the underspecified model severely underperforms in predicting returns relative to the benchmark model. The increase in the mean *U* statistic for the restricted model is associated with a decrease in the mean $\bar{R}^2$ from 0.504 for the benchmark model to 0.142 for the restricted model. This is in line with Frank's (2009: 58) hypothesis that higher *U* statistics imply poor model performance. The upward shifts in the minimum and maximum values confirm the underperformance of the restricted model. The minimum *U* statistic of 0.547, for the food producers sectors, is substantially greater than that of 0.120 for the mining sector in Panel B of Table 8.4. The maximum *U* statistic of 0.767 for the fixed telecommunications sector is greater that of 0.602 for this sector in the benchmark model (Panel B of Table 8.4.).

The bias and variance proportions, U_{BIAS} and U_{VAR}, provide further insight into the consequences of underspecification. The overall bias proportion of 0.002355 in Panel B of Table 9.4. is significantly higher than that of 0.001036 for the benchmark specification in Panel B of Table 8.4. The sector with the lowest bias proportion is now the pharmaceutical and biotechnology sector with a bias proportion of almost zero. However, the maximum value has increased to 0.020372 for the software and computer services sector, which has a bias proportion of 0.017506 in the benchmark model. Although the mean bias proportion more than doubles, it remains below a level of 0.1, which would be deemed as concerning (Brooks & Tsolacos, 2010: 272). Nevertheless, the increase in the overall bias proportion indicates that the restricted model is prone to greater systematic errors in predictions. The mean variance proportion component of the Theil *U* statistic significantly increases from 0.178215 for the benchmark specification in Panel B of Table 8.4. to 0.412575 for the restricted specification in Panel B of Table 9.4. This represents a more than two-fold increase in the magnitude of the mean measure of the variance proportion. The minimum and maximum values also reflect this increase in the variance proportion and show an upward shift. The mining sector has the lowest variance proportion of 0.022184 for the benchmark specification whereas the lowest variance proportion is now 0.287454 for the industrial transport sector. The computer software and services sector, which has the highest variance proportion of 0.509499 for the benchmark model, now has a variance proportion of 0.613429.

The bias proportion is indicative of systematic over- or underprediction and the variance proportion is an indicator of the assumed linear factor model's ability to replicate the second moment of realised returns, the variance. Elkhafif (1996: 97) states that the variance proportion measures the degree to which the predicted values replicate the actual variability of a series and indicates the model's ability to replicate turning points. In the present case, the mean of the variance proportion increases significantly. Unlike the bias proportion for which values remain below 0.1, the increase in the overall variance proportion suggests that the model loses its ability to accurately replicate the factor of interest, namely returns, and turning points in the return series. The inference that can be drawn so far is that the restricted linear factor model suffers from a reduced ability to fulfil its core function, namely that of modelling price changes accurately.

The mean covariance proportion decreases significantly from 0.807705 for the benchmark specification to 0.585071 for the restricted specification. The sectors that now have the highest and lowest covariance proportions are the industrial transport sector with a covariance proportion of 0.711488 and the software and computer services sector with a covariance proportion of 0.366200. This reflects a downward shift in the minimum and maximum covariance proportions relative to the benchmark model. For the benchmark model, the mining sector is associated with the highest covariance proportion of 0.976984 whereas the lowest covariance proportion of 0.472995 is reported for the software and computer services sector. The results presented for the restricted model suggest that factor omission introduces a greater level of systematic prediction error into the model. Consequently, a greater proportion of the prediction error is now attributable to the omitted factors and the model itself than random events that are series specific (Elkhafif, 1996: 97; Brooks & Tsolacos (2010: 272)).

The results presented in this section suggest that factor omission will result in significantly larger prediction errors. As argued by Chang (1991: 387), this will negatively impact the power of statistical tests. This, together with the upward bias in the standard errors and residual variance reported in Section 9.5., can explain the observed understatement of the importance of macroeconomic factors that feature in the linear factor model (Section 9.3.1.). Factor omission also adversely impacts the ability of a specification to predict returns accurately and the evidence presented in this section supports this. Underspecified models will underperform in predicting the mean and the variability of the return series. Finally, factor omission will introduce greater systematic error into the model. A greater proportion of prediction error will be directly attributable to the model specification.

9.7. FACTOR OMISSION

The results of the LR test for omitted factors are summarised in Panel A of Table 9.5. and confirm that the restricted specification is underspecified. The null hypothesis of $M\varepsilon_t$ being insignificant is rejected for all series in the sample indicating that the residual market factor proxies for omitted factors. However, it appears that $M\varepsilon_t$ is not a proxy for all omitted factors. The null hypothesis that $IM\varepsilon_t$ is insignificant is rejected for 10 out of 26 industrial sectors modelled using the restricted specification. This suggests that a second residual market factor, $IM\varepsilon_t$, reflects information that is not captured by the macroeconomic factor set in equation (9.1) and also by $M\varepsilon_t$. However, the relevance of this factor is restricted to less than half of the series in the sample. This suggests that in the present context, the impact of $IM\varepsilon_t$ is limited to a subset of sectors. More concerningly, the null hypothesis that the factors jointly comprising the factor analytic augmentation, f_{1t} and f_{2t}, are insignificant is rejected for all sectors, with the exception of the software and computer services sector (see Table A1.2. in Appendix A). This points towards the presence of relevant information for returns in the residual correlation matrix of the unrestricted model (equation (8.1)), which incorporates the macroeconomic factor set and the two residual market factors. Importantly, this again confirms that the restricted specification omits relevant factors and is therefore underspecified (Section 6.4.8.).

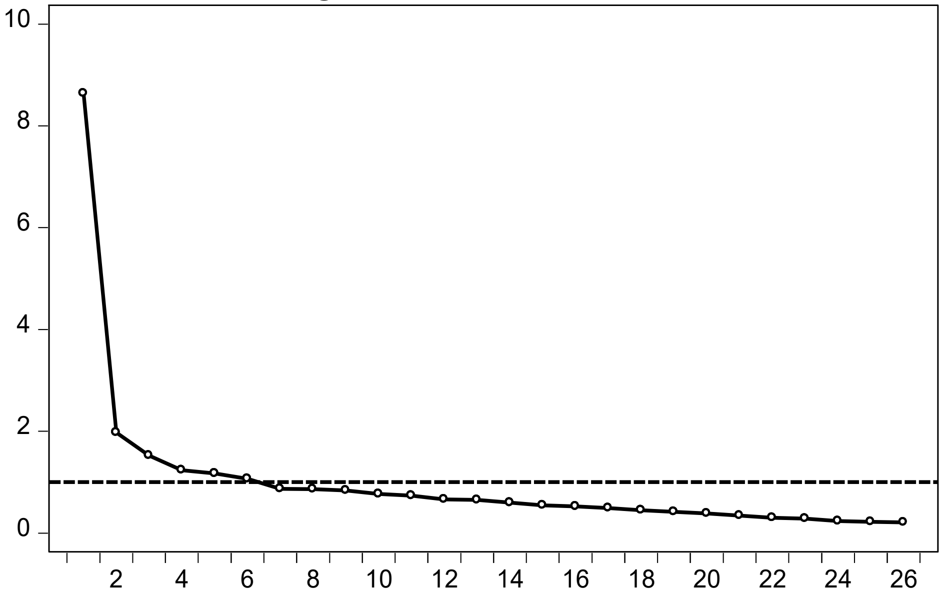
Table 9.5: Likelihood Ratio Test And Factor Analysis Summary For The Restricted Model

Panel A: Likelihood Ratio Test for Omitted Factors		
Omitted Factor(s)	**Mean LR Statistic**	**Total Sig.**
$M\varepsilon_t$	50.038	26/26
$IM\varepsilon_t$	3.182	10/26
f_{1t}, f_{2t}	49.014	25/26
Panel B: Full Period Factor Analysis		
Extracted	**Mean Communality**	**Mean Uniqueness**
3	0.399▲	0.601
Panel C: Subperiod Factor Analysis		
Period: 2001M01 to 2008M12		
3	0.461▲	0.539▼
Period: 2009M01 to 2016M12		
2	0.339▲	0.661▼

Notes:
In Panel A, the Mean LR Statistic is the mean of the LR test statistics from the Likelihood Ratio test for omitted factors. Total Sig. is the number of outcomes rejecting the null hypothesis that a given factor or set of factors has not been omitted. Significance is recorded at the 10% level of significance. In Panel B and Panel C, Mean Communality is the mean proportion of common variance explained across return series by common factors extracted on the basis of the MAP test. Mean Uniqueness is the mean proportion of variance across return series attributable to the return series themselves and not systematic factors. Across panels, the first ▲ or ▼ symbol indicates that a value is larger or smaller relative to that observed for the benchmark model.

The scree plot in Figure 9.1. indicates that there are two common factors in the residual correlation matrix of the restricted model. This is similar to the results of the scree test applied in Section 7.2.

Figure 9.1: Scree Plot Of Eigenvalues For Residuals Of The Restricted Model



In contrast to the results of the scree test and as with the actual return series in Section 7.2., the MAP test indicates that there are three common factors in the residuals of the restricted specification (Panel B of Table 9.5.; Section 7.2.). An examination of the mean communality and uniqueness associated with the factors derived from the residual correlation matrix of the restricted model suggests that these differ from the factors derived from the benchmark model residuals. Whereas the single factor derived from the benchmark model residuals has a relatively small mean communality of 0.066, as reported in Panel A of Table 8.5., the three factors derived from the residuals of the restricted model have a mean communality of 0.399, as evident from Panel B of Table 9.5. This suggests that an additional 40% of the variation in industrial sector returns is explained by omitted factors. The mean uniqueness of 0.601 is below that of the benchmark model of 0.934. All sectors have communalities of above 0.15 (unreported) and it is therefore difficult to argue that these factors arise from strong interdependence between a limited number of sectors, as argued in Section 8.7. for the residuals of the benchmark model. These results suggest that a greater proportion of the shared variance in returns relative to the benchmark model is attributable to omitted systematic factors than to unique factors in the residual series as would be expected if all relevant systematic factors have been taken into consideration. It is difficult to argue that the three factors extracted from the residual correlation matrix of the restricted model are pseudofactors, as appears to be the case for the factors extracted from the residuals of the benchmark model in Section 8.7.

Results of the MAP test for the 2001M01 to 2008M12 period indicate that there are three factors in the residual correlation matrix whereas there are two factors in the residual correlation matrix for the 2009M01 to 2016M12 period (Panel C of Table 9.5.). The mean communalities for the 2001M01 to 2008M12 and the 2009M01 to 2016M12 periods are 0.461 and 0.339 respectively whereas the respective mean uniqueness measures are 0.539 and 0.661. That factors are extracted for the 2001M01 to 2008M12 period is in contrast to the results in Section 8.7. for the benchmark specification. No factors are identified over this period. For the 2009M01 to 2016M12 period, a single factor is extracted from the residuals of the benchmark specification. The mean communality of 0.339 for the restricted model is also substantially greater for this period, relative to the mean communality of 0.097 for this period for the benchmark model.

The discrepancy between the number of factors (three and two) and the mean communalities across the two subperiods (0.461 and 0.339) may be attributable to changes in the structure of the return generating process. Panetta (2002: 443-444) states that the nature of the relationships between stock returns and macroeconomic factors may change over time. Certain macroeconomic factors may be more adept at explaining returns during certain periods as opposed to others. Therefore, it is plausible that the return generating process changes over time and this is a potential reason for underspecification, namely the inability of the same macroeconomic factors to sufficiently and adequately explain returns over time (McQueen & Roley, 1993). This may explain the discrepancies between the number of factors and the shared proportion of variance that is explained by the factors extracted from the residuals of the restricted model over the two subperiods. Specifically, it may be that the macroeconomic factor set performs poorly in the first subperiod but better in the second subperiod in explaining returns. Consequently, the residuals of the restricted model in the second subperiod reflect fewer common factors and lower interdependence, as suggested by lower mean communality. Importantly, what is certain is that the factors extracted from the residual correlation matrix over the two subperiods are not transitory.

The results of the factor analysis presented in this section point towards the presence of omitted pervasive influences, which are relegated to the residuals. These factors appear to be systematic in nature, general and non-transitory. In other words, these factors cannot be classified as pseudofactors and as such, the restricted specification is underspecified. This is supported by the results of the LR test for omitted factors reported in Panel A of Table 9.5.

9.8. THE RESIDUAL CORRELATION MATRIX

That the restricted model is underspecified is confirmed by the LR test for omitted factors and the results of factor analysis reported in Table 9.5. These results imply that the correlation matrix derived from the residuals of the restricted model reflects omitted factors and differs from that of the benchmark model. Figure 9.2. and Table 9.6. provide further evidence that this is the case.

Figure 9.2: Histogram Of Restricted Model Residual Correlation Coefficients

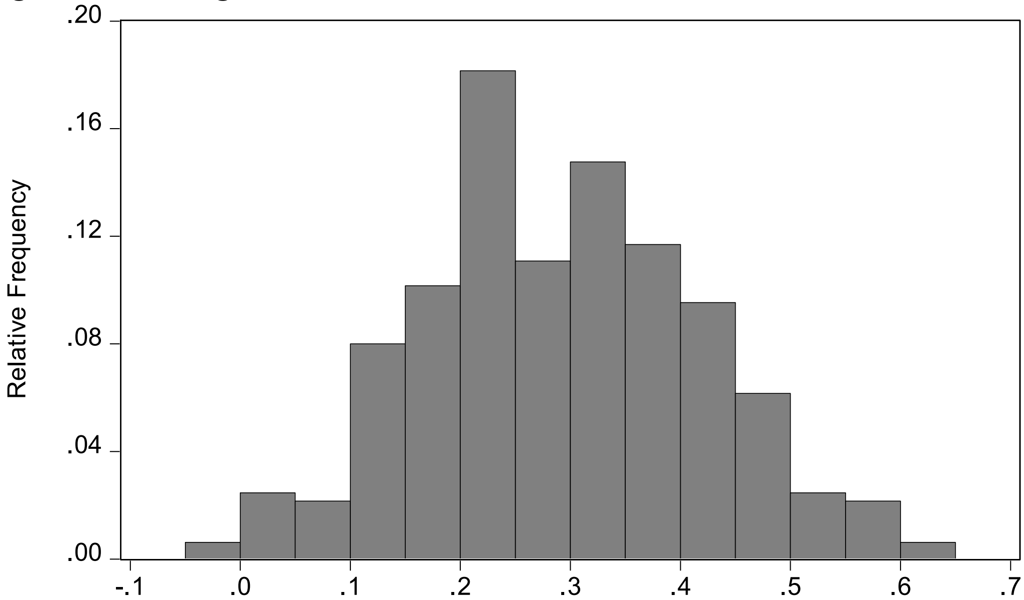


Table 9.6. Distribution Summary Of The Restricted Model Residual Correlation Matrix

Bin	Frequency	Relative Frequency	Cumulative Frequency
$-0.5 < \rho_{ij} \leq -0.4$	0	0.000%	0.000%
$-0.4 < \rho_{ij} \leq -0.3$	0	0.000%	0.000%
$-0.3 < \rho_{ij} \leq -0.2$	0	0.000%	0.000%
$-0.2 < \rho_{ij} \leq -0.1$	0	0.000%	0.000%
$-0.1 < \rho_{ij} \leq 0$	2	0.615%	0.615%
$0 < \rho_{ij} \leq 0.1$	15	4.615%	5.231%
$0.1 < \rho_{ij} \leq 0.2$	59	18.154%	23.385%
$0.2 < \rho_{ij} \leq 0.3$	95	29.231%	52.615%
$0.3 < \rho_{ij} \leq 0.4$	86	26.462%	79.077%
$0.4 < \rho_{ij} \leq 0.5$	51	15.692%	94.769%
$0.5 < \rho_{ij} \leq 0.6$	15	4.615%	99.385%
$0.6 < \rho_{ij} \leq 0.7$	2	0.615%	100.000%
Total	**325**	**100%**	**100.000%**
Mean	0.291***▲***		
Minimum	-0.003		
Maximum	0.615		

Table 9.6. Distribution Summary Of The Restricted Model Residual Correlation Matrix (Continued…)

Notes:
The asterisks, ***, ** and *, indicate statistical significance at the respective 1%, 5% and 10% levels of significance. The *t*-test is applied to test the hypothesis that the mean of the correlation coefficients does not differ significantly from zero. The Wilcoxon matched-pairs signed-rank test is applied as a confirmatory test and the superscript "W" indicates that the Wilcoxon matched-pairs signed-rank test contradicts the results of the paired-sample *t*-test. . Bin represents ranges of correlation coefficients and frequency reports the number of correlation coefficients that fall within each range. Relative Frequency is the percentage of correlation coefficients that fall within the respective ranges. Cumulative Frequency is the running total of all previous relative frequencies. Mean is the mean value of the correlation coefficients in the correlation matrix and the Minimum and Maximum are the lowest and largest correlation coefficients observed. Across Panels, the first ▲ or ▼ symbol indicates that a value is smaller or larger relative to that observed for the benchmark model. Accompanying asterisks, if present, indicate that differences are statistically significant.

The histogram of the residual correlation coefficients in Figure 9.2. contrasts with that in Figure 8.1. It is immediately evident that 99.385% (323 of 325) of the resultant correlation coefficients are positive. Only two correlation coefficients are negative whereas the benchmark residual correlation matrix exhibits a spread of both negative and positive correlation coefficients. In fact, 76.615% (249 out of 325) of coefficients are greater than 0.2 in magnitude. In comparison, only 1.538% (5 out of 325) of correlation coefficients in the benchmark model residual correlation matrix are greater than 0.2 (not absolute). While 71.629% (233 out of 325) of the coefficients in the residual correlation matrix derived from the benchmark model are smaller than 0.12 in absolute magnitude, the magnitude at which coefficients tend to be statistically insignificant, only 7.077% (23 of 325) of correlation coefficients fall within this range for the residual correlation matrix derived from the restricted specification. This suggests that while most of the residual correlation coefficients in the benchmark model residual correlation matrix are statistically insignificant, most of the residual correlation coefficients derived from the restricted model are statistically significant. These findings are indicative of widespread residual interdependence for the restricted model, which can be attributed to the presence of omitted common factors, as supported by the findings in Section 9.7. (King, 1966: 142). Such findings are expected if a model is underspecified.

The mean level of residual correlation is 0.291 and is significantly different from zero. This is indicative of the tendency of the residual correlation summarised in Table 9.6. to be positive and larger in magnitude relative to that of the benchmark model (which is -0.024). Importantly, this figure is not substantially lower than that of the mean correlation of actual returns of 0.375 (Table 7.1.). This suggests that the residual correlation structure of the restricted model more closely resembles that of the actual return series than that of the benchmark model. Differences between the mean residual correlation coefficients of the benchmark and restricted models are statistically significant, confirming that overall, the level of residual correlation for the restricted model is higher than that of the benchmark model.

The extremes of the estimated residual correlation coefficients also reflect the shift from a correlation structure that exhibits a spread of both positive and negative correlation coefficients , namely correlation coefficients that generally do not conform to a set pattern of interdependence, to one that is indicative of overwhelmingly positive interdependence. The estimated residual correlation coefficients range between -0.003 and 0.615. This is in contrast to a more balanced range of correlation coefficients ranging between -0.320 and 0.396 for the benchmark model. Notably, the extremes for the correlation coefficients of the restricted model are closer to those of the actual returns, which range between 0.048 to 0.673 (Table 7.1.). Comparisons of the minimum and maximum values, as with the mean level of correlation, point towards a residual interdependence structure that does not differ significantly from that of the actual return series. This suggests that the macroeconomic factors, by themselves in the restricted model, perform poorly in explaining the co-movement between stocks (Elton *et al.,* 2014: 157). Such a result is perhaps to be expected, based upon prior findings; Connor (1995: 44) reports that (five) macroeconomic factors have low explanatory power for a sample of US stocks.

The difference between the residual correlation matrix in Table 9.7. and the residual correlation matrix in Table 8.7. is readily discernible. While Table 8.7. does not exhibit systematic patterns of significant correlation and correlation of a consistent direction (either positive or negative), the correlation coefficients in Table 9.7. are statistically significant and positive, with very few exceptions (unshaded). In fact, out of a total of 325 estimated correlation coefficients, only 22 are not statistically significant and a single coefficient is negative (and also statistically insignificant). This overwhelmingly significant and positive correlation structure confirms that the restricted model, specifically the macroeconomic factor set that comprises this model, fails to explain much of the co-movement that is associated with common factors that are now relegated to the residuals. Positive significant correlation is also observed within economic sectors and between economic sectors indicating that economic sectors and industrial sectors move together in response to omitted common factors.

Table 9.7: Correlation Matrix Of The Restricted Model Residuals

	J135	J173	J175	J177	J235	J272	J273	J275	J277	J279	J335	J353	J357	J453	J457	J533	J537	J555	J575	J653	J835	J853	J857	J877	J898	J953
J135	1.000																									
J173	0.281	1.000																								
J175	0.194	0.134																								
J177	0.216	0.341	0.550	1.000																						
J235	0.414	0.202	0.332	0.223	1.000																					
J272	0.305	0.264	0.182	0.244	0.399	1.000																				
J273	0.457	0.252	0.217	0.234	0.504	0.415	1.000																			
J275	0.350	0.275	0.247	0.273	0.312	0.203	0.471	1.000																		
J277	0.462	0.242	0.215	0.254	0.400	0.406	0.432	0.414	1.000	0.486																
J279	0.365	0.304	0.113	0.124	0.432	0.503	0.404	0.193	0.486	1.000																
J335	0.350	0.170	0.184	0.128	0.144	0.193	0.261	0.254	0.294	0.216	1.000															
J353	0.265	0.380	0.087	0.253	0.104	0.408	0.341	0.246	0.236	0.215	0.168	1.000														
J357	0.306	0.125	0.200	0.222	0.286	0.461	0.415	0.206	0.448	0.392	0.226	0.353	1.000													
J453	0.349	0.201	0.093	0.166	0.283	0.385	0.329	0.212	0.409	0.307	0.155	0.321	0.327	1.000												
J457	0.228	0.043	0.128	0.046	0.188	0.184	0.277	0.121	0.232	0.247	0.205	0.149	0.279	0.326	1.000											
J533	0.202	0.014	0.105	0.012	0.265	0.453	0.218	0.098	0.381	0.314	0.137	0.140	0.416	0.302	0.273	1.000										
J537	0.301	0.101	0.083	0.041	0.348	0.552	0.444	0.234	0.565	0.525	0.280	0.214	0.514	0.400	0.423	0.580	1.000									
J555	0.340	0.169	0.000	0.131	0.206	0.463	0.299	0.168	0.434	0.430	0.173	0.317	0.287	0.306	0.234	0.227	0.458	1.000								
J575	0.423	0.250	0.148	0.163	0.401	0.381	0.358	0.394	0.520	0.480	0.249	0.328	0.402	0.364	0.230	0.219	0.405	0.359	1.000							
J653	0.212	0.027	-0.003	0.138	0.114	0.261	0.201	0.084	0.274	0.256	0.142	0.021	0.234	0.110	0.233	0.194	0.332	0.366	0.221	1.000						
J835	0.244	0.139	0.159	0.131	0.308	0.558	0.350	0.228	0.462	0.412	0.127	0.207	0.404	0.304	0.298	0.396	0.614	0.372	0.384	0.259	1.000					
J853	0.272	0.084	0.223	0.175	0.267	0.334	0.307	0.190	0.419	0.224	0.173	0.247	0.403	0.355	0.198	0.331	0.329	0.162	0.299	0.165	0.311	1.000				
J857	0.373	0.319	0.128	0.177	0.230	0.456	0.362	0.305	0.405	0.398	0.206	0.374	0.392	0.332	0.281	0.332	0.472	0.416	0.390	0.246	0.615	0.346	1.000			
J877	0.383	0.226	0.021	0.169	0.301	0.530	0.455	0.166	0.382	0.457	0.190	0.401	0.404	0.212	0.239	0.374	0.548	0.495	0.418	0.320	0.594	0.226	0.584	1.000		
J898	0.359	0.319	0.065	0.241	0.267	0.428	0.338	0.160	0.358	0.383	0.252	0.500	0.377	0.286	0.240	0.251	0.381	0.406	0.385	0.230	0.311	0.321	0.513	0.481	1.000	
J953	0.201	0.208	0.133	0.273	0.165	0.338	0.302	0.130	0.335	0.280	0.194	0.219	0.278	0.185	0.236	0.187	0.315	0.476	0.345	0.331	0.361	0.321	0.494	0.493	0.373	1.000

Table 9.8: Tests Of Matrix Equality For The Restricted Model

Hypothesis	χ^2 Statistic	Reject
$R_{26} = A_{26}$	102.768	Fail
$R_{26} = I_{26}$	6258.94***	Reject
$R_{26} = B_{26}$	838.788***	Reject

Notes:
The asterisks, ***, ** and *, indicate statistical significance at the respective 1%, 5% and 10% levels of significance. Hypothesis is the hypothesis that is being tested relating to the equality of two matrices. χ^2 Statistic is the resultant test statistic for the Jennrich test and Reject indicates whether the null hypothesis of equality between two matrices is rejected. B_{26} denotes the residual correlation matrix derived from the benchmark model. A_{26} denotes the residual correlation matrix of the actual return series. I_{26} denotes the identity matrix. R_{26} denotes the residual correlation matrix derived from the restricted model. B_{26} denotes the residual correlation matrix derived from the benchmark model. A_{26} denotes the residual correlation matrix of the actual return series. I_{26} denotes the identity matrix. R_{26} denotes the residual correlation matrix derived from the restricted model.

The Jennrich (1970) test is applied to determine whether the residual correlation matrix derived from the restricted specification, R_{26}, differs significantly from that of the actual return series, A_{26}. Results indicate that the null hypothesis of the equality of matrices cannot be rejected. This is a severe indictment of the ability of macroeconomic factors to account for the co-movement in industrial sector returns to such an extent that much of the original correlation structure remains intact. This also constitutes further evidence that there are other factors that have been omitted from the characterisation of the return generating process (Elton & Gruber, 1988: 31; Elton *et al.,* 2014: 157).

The equality between R_{26} and I_{26}, the identify matrix, and between R_{26} and B_{26}, the benchmark model residual correlation matrix, is also tested. Expectedly, given the results reported in Table 8.8. for the test of equality between B_{26} and I_{26}, the null hypothesis of equality between R_{26} and I_{26} is also rejected. The result is highly significant; the χ^2 statistic is 6258.94 whereas the χ^2 statistic for the test of equality between B_{26} and I_{26} is 826.360 in Table 8.8. The null hypothesis of the equality between R_{26} and B_{26} is also rejected. This indicates that there is a statistically significant difference between the residual correlation structure of the restricted model and that of the benchmark model. Furthermore, this indicates that the additional factors incorporated into the benchmark model, namely the two residual market factors and the factor scores, account for a significant proportion of the co-movement in returns attributable to pervasive factors. This co-movement is not captured by the macroeconomic factor set alone.

These findings, together with the general findings of Section 9.7. and the analysis relating to the distribution of the correlation coefficients summarised in Table 9.6, point towards a correlation structure that reflects high levels of interdependence. The tests of equality between R_{26} and A_{26} are particularly revealing and concerning. The assumption of unrelated residuals should hold if a specification is valid and no strong factors should be reflected in the residuals (Meyers, 1973: 698). The results presented in this section suggest that this assumption is widely violated and the residuals reflect a strong interdependence structure, attributable to the presence of omitted factors (Table 9.5.).

9.9. CHAPTER SUMMARY AND CONCLUSION

The restricted model comprises seven macroeconomic factors, which that are shown to be proxies for the pervasive influences in stock returns (equation (7.1)) in Section 7.4.). It is implicitly assumed that this model is underspecified as it omits the residual market factors and the factor analytic augmentation.

Notable differences are observed between this model and the benchmark model across a number of aspects. Fewer coefficients on the macroeconomic factors are statistically significant and the importance of factors is understated. Importantly, two factors, $LEAD_{t-1}$ and $USD\varepsilon_t$, no longer appear to be systematic. This erroneously suggests that these factors should be excluded from further tests of the APT, if such tests were to be carried out (Section 9.3.1.). Coefficient bias increases overall as indicated by larger differences between ML coefficient estimates for the restricted model and least squares coefficient estimates for the benchmark model. This potentially exacerbates an errors-in-variable problem (Section 9.3.2.). The explanatory power of the restricted model is significantly lower than that of the benchmark model suggesting that macroeconomic factors, by themselves, yield a poor description of the return generating process. The mean AIC and BIC statistics are significantly higher. This indicates a deterioration in the predictive accuracy of the model and a greater deviation from the true return generating process (Section 9.3.3.).

Residual diagnostics reported in Section 9.4. indicate that the exclusion of the abovementioned factors translates into a greater number of instances of serial residual correlation. In Section 9.5., it is shown that residual variance is significantly higher in general and in comparisons across individual sectors. This presents a potential explanation for the understatement and misidentification of factors in the linear factor model. The complexity of the conditional variance structures increases; the variance for a greater number of sectors is modelled as a GARCH(1,1) process. Changes in the structure of the conditional variance potentially explain the overall increases in the differences between ML and least squares coefficients (Panel A of Table 9.1.).

The mean errors of the restricted model significantly differ from zero and are greater (in absolute terms) than those of the benchmark model, which are statistically insignificant. Larger Theil *U* statistics are indicative of a deterioration in predictive performance and the mean bias and variance proportions suggest that the model underperforms the benchmark model in predicting actual mean values and the variance of the series. A greater proportion of prediction error is now attributable to the representation of the return generating process itself in the form of the specification of the linear factor model (Section 9.6.).

The LR test for omitted factors indicates that the restricted model systematically and widely omits factors reflected by the residual market factor and the factor analytic augmentation. Factor analysis of the resultant correlation matrix in Section 9.7. suggests that pervasive influences remain in the residuals and are not captured by the macroeconomic factors in the restricted model. Concerningly, an additional 40% of variation may be explained by factors that are not reflected in the model. Finally, the analysis of the residual correlation matrix in Section 9.8. and the Jennrich (1970) test indicate that the residual correlation matrix is equal to the correlation matrix of the actual return series.

In summary, macroeconomic factors appear to be poor proxies for pervasive influences in returns and do not capture most of the interdependence in returns attributable to the influence of common factors. The assumption of uncorrelated residuals across sectors is widely violated. This challenges the validity of a linear factor model that solely incorporates macroeconomic factors.

The use of a residual market factor is an approach commonly applied in APT literature to resolve underspecification in linear factor models that rely on macroeconomic factors to proxy for the underlying pervasive influences in returns. If the residual market factor is an adequate proxy for omitted factors, then the inclusion of this factor should translate into results that are comparable to those of the benchmark model. If orthogonal to the macroeconomic factor set and the conventional residual market factor, a second residual market factor should be irrelevant. This efficacy of the residual market factor as a proxy for omitted factors and the relevance of a second residual market factor are investigated in Chapter 10.

CHAPTER 10

UNDERSPECIFICATION, THE RESIDUAL MARKET FACTORS AND THE LINEAR FACTOR MODEL

10.1. INTRODUCTION

The results in Chapter 9 indicate that macroeconomic factors in the linear factor model by themselves produce a poor approximation of the return generating process. For this reason, a conventional residual market factor is incorporated into the linear factor model to function as proxy for the remaining pervasive factors that influence stock returns. This chapter investigates whether the conventional residual market factor is an adequate and sufficient proxy for omitted factors in the macroeconomic linear factor model. The role of a second residual market factor is also considered. The residual market factors are derived from returns on the JSE All Share Index, the South African stock market aggregate, and returns on the MSCI World Market Index, a commonly used proxy for international/global influences in stock returns (Section 4.5.).

It is possible that the residual market factors fail to fully account for omitted factors and evidence suggestive of this is provided by the benchmark model. The benchmark specification in Chapter 8 comprises the macroeconomic factor set, the two residual market factors and a factor analytic augmentation. The factor analytic augmentation comprises factor scores derived from the residuals of the unrestricted specification (equation (8.1)). The extraction of factors from the residuals of the unrestricted specification and their widespread significance in the benchmark specification (Panel A of Table 8.1.) implies that the residual market factors fail to account for omitted factors. The widespread statistical significance of $IM\varepsilon_t$, the international residual market factor that is orthogonal to all other factors in the benchmark specification and the conventional residual market factor, suggests that the macroeconomic factor set and $M\varepsilon_t$, fail to account for remaining pervasive influences. Nonetheless, it is possible that a single residual market factor or perhaps even two residual market factors yield an adequate description of the linear factor model that is not impacted materially by underspecification.

This chapter investigates whether the residual market factor is an adequate proxy for omitted factors and whether its inclusion alongside macroeconomic factors contributes to an adequate specification of the linear factor model. The premise of this analysis is that if the residual market factor is an adequate proxy for omitted factors that resolves underspecification, then $IM\varepsilon_t$ should be redundant and the unrestricted market model should approximate the fully-specified benchmark model.

This chapter follows the same structure as that of Chapter 8 and Chapter 9. Section 10.2. outlines the specifications of the unrestricted market model and the unrestricted model and Section 10.3. provides a general overview of the results and associated comparisons. Section 10.4. reports the results of the diagnostic tests and robustness checks. Section 10.5. investigates the residual variance and the conditional variance structures. Predictive ability is investigated in Section 10.6. and factor analysis is applied to the resultant correlation matrices in Section 10.7. The structure of the residual correlation matrices is directly considered in Section 10.8. Throughout Chapter 10, comparisons are made to the results of the restricted model and the benchmark model. This is to establish whether a single residual market factor or two residual market factors produce an improvement in the representation of the return generating process and whether a linear factor model combining macroeconomic factors and the residual market factor or factors approximates the benchmark model.

10.2. UNRESTRICTED MODEL SPECIFICATIONS

To investigate the ability of the conventional residual market factor and an international residual market factor to proxy for omitted factors, two unrestricted specifications are estimated:

$$\begin{aligned} R_{it} &= \alpha + b_{iBP}BP_{t-1} + b_{iLEAD}LEAD_{t-1} + b_{iBUS}BUS_t + b_{iUSD\varepsilon}USD\varepsilon_t + b_{iMET}MET_t + b_{iLTY}LTY_t \\ &+ b_{iTLI}TLI\varepsilon_t + b_{iM\varepsilon}M\varepsilon_t + \varepsilon_{it} \end{aligned} \tag{10.1}$$

$$\begin{aligned} R_{it} &= \alpha + b_{iBP}BP_{t-1} + b_{iLEAD}LEAD_{t-1} + b_{iBUS}BUS_t + b_{iUSD\varepsilon}USD\varepsilon_t + b_{iMET}MET_t + b_{iLTY}LTY_t \\ &+ b_{iTLI}TLI\varepsilon_t + b_{iM\varepsilon}M\varepsilon_t + b_{iIM\varepsilon}IM\varepsilon_t + \varepsilon_{it} \end{aligned} \tag{10.2}$$

where in equations (10.1) and (10.2), all notation remains as in equations (8.1) and (9.1). However, the unrestricted market model denoted by equation (10.1) now incorporates the domestic residual market factor, $M\varepsilon_t$ derived from returns on the JSE All Share Index. Equation (10.2) is the unrestricted model that comprises the macroeconomic factors, $M\varepsilon_t$ and the international residual market factor, $IM\varepsilon_t$. As with the benchmark specification in Chapter 8 and the restricted specification in Chapter 9, equations (10.1) and (10.2) are estimated using ML estimation. The conditional variance of each series is initially assumed to follow an ARCH*(p)* or GARCH(*p,q)* process identical to that of the corresponding benchmark specification. Subsequently, the number of ARCH and/or GARCH parameters is increased accordingly until the residuals are free of non-linear dependence and heteroscedasticity. If the residual market factor in equation (10.1) is an adequate proxy for omitted factors, then the benchmark and unrestricted market model should be comparable across the different aspects considered. Furthermore, $IM\varepsilon_t$ in equation (10.2) should not be broadly statistically significant nor should the unrestricted model present a significant improvement over the unrestricted market model.

10.3. MODEL OVERVIEW AND COMPARISONS

10.3.1. Macroeconomic Factor Significance Comparisons

The abridged results of both unrestricted specifications are reported in Table 10.1. The investigation into ability of the two residual market factors to proxy for omitted factors begins with a consideration of the number of statistically significant coefficients associated with the macroeconomic factors. The findings in Chapter 9 show that underspecification translates into an understatement of the significance of factors. This is attributable to inflated residual variance, resulting in a loss of efficiency and misleading inferences (Section 9.3.; Section 9.5.). Moreover, the results of the LR test, the analysis of the pairwise residual correlation coefficients and factor analysis conducted on the residual correlation matrix of the restricted model points towards the existence of omitted systematic factors (Section 9.7.; Section 9.8.).

If the residual market factor is an adequate proxy for omitted factors, the number of statistically significant macroeconomic factor coefficients should approach that of the benchmark model. For the unrestricted market model, 105 of the 182 (57.692%) estimated coefficients for the macroeconomic factors are statistically significant whereas for the unrestricted model, 109 of the 182 (59.890%) estimated coefficients are statistically significant. For both specifications, this is an improvement relative to the restricted model. The number of statistically significant coefficients for both specifications is closer to that of the benchmark specification, for which 119 of the 182 (65.38%) estimated coefficients are statistically significant. For the unrestricted market model, an additional 12 coefficients (an additional 12.093%) are statistically significant relative to the restricted model discussed in

Chapter 9. For the unrestricted model, an additional 16 coefficients (an additional 17.20%) are statistically significant.[146]

Although the inclusion of the domestic residual market factor and the international residual market factor appears to progressively alleviate the understatement of factor significance, it does not completely eliminate understatement. When coefficients on the residual market factor are also considered in the unrestricted market model, the number of statistically significant coefficients is 131 out of 208 (62.981%) estimated coefficients. This number is still below the 145 out of 208 (69.71%) statistically significant coefficients associated with these factors in the benchmark model. The number of statistically significant coefficients for the unrestricted model that incorporates the international residual market factor is 149 out of 234 (63.367%). In this model, fewer coefficients on $IM\varepsilon_t$ are statistically significant relative to the benchmark model (-2). Although the inclusion of the residual market factor in the unrestricted models appears to improve inference, possibly following a reduction in the magnitude of the residual variance and standard errors (discussed in Section 10.5.), the specification still appears to understate the importance of factors. It appears that the contribution of $IM\varepsilon_t$ to alleviating understatement is marginal although this factor is statistically significant in 14 instances in the unrestricted model. This suggests that $IM\varepsilon_t$ reflects additional information that is not reflected by $M\varepsilon_t$.

[146] Reported are the net numbers of statistically significant coefficient estimates, following the inclusion of the residual market factors. The net amounts take into account overstatement and understatement. For example, the number of significant estimates of MET_t is higher relative to that of the benchmark specification (+1 and +2 for the respective unrestricted specifications). Also, returns on the industrial metals and mining sector exhibit a positive and statistically relationship with LTY_t whereas in the benchmark model, this relationship is negative and statistically insignificant. The robustness of the results is investigated in Section 10.4. Given that these inconsistencies are not widespread and do not impact the overall results of this chapter, these isolated discrepancies are not investigated beyond the robustness checks in Section 10.4.

Table 10.1: Summary Of Unrestricted Model Results

Factor	Mean Coeff. Std Error |Z-score|	Mean LS Co. | Diff. |	$b_{ik} > 0$	$b_{ik} = 0$	$b_{ik} < 0$	Total Sig.	ΔSig	Mean Coeff Std Error |Z-score|	Mean LS Co. | Diff. |	$b_{ik} > 0$	$b_{ik} = 0$	$b_{ik} < 0$	Total Sig.	ΔSig
Panel A: Coefficient And Significance Summary														
Unrestricted Market Model								**Unrestricted Model**						
Intercept	0.007▲**▼*** (0.004)▲▼ [1.955]▼▲	0.006 0.001**▲▼	16	10	-	16	+3	0.007▲*▼** (0.004)▲▼ [1.881]▼▲	0.006 0.001**▲▼	15	11	-	15	+2
BP_{t-1}	0.035▼▲** (0.037)▲▼ [1.147]▼▲	0.038 0.003▲▼	7	19	-	7	-3	0.035▼▲*** (0.037)▲▼ [1.187]▼▲	0.038 0.003▲▼	9	17	-	9	-1
$LEAD_{t-1}$	0.925▲▲** (0.488)▲▼ [2.026]▼▲	0.883 0.042▲▼	13	13	-	13	-4	0.921▲▲** (0.486)▲▼ [2.038]▼▲	0.883 0.038▲▼	13	13	-	13	-4
BUS_t	0.070▼*▼* (0.045)▲▼ [1.782]▼▲	0.079 0.009▲▲	15	11	-	15	-3	0.072▼* ▼W (0.044)▲▼ [1.824]▼▲	0.079 0.007▲▲	15	11	-	15	-3
$USD\varepsilon_t$	-0.163▲▲ (0.131)▲▼ [1.683]▼▲	-0.174 0.010▲▲	1	13	12	13	-3	-0.167▲▼ (0.130)▲▼ [1.743]▼▲	-0.174 0.007▲▼	1	12	13	14	-2
MET_t	0.160▲▲ (0.087)▲▼ [2.028]▼▲	0.173 0.013▼▼	14	11	1	15	+1	0.158▲▲ (0.086)▲▼ [2.025]▼▲	0.173 0.015▼▼	15	10	1	16	+2
LTY_t	-3.658▲▲** (1.299)▲▼ [3.344]▼▲	-3.727 0.069▼▼	1	6	19	20	-	-3.666▲▲** (1.341)▲▼ [3.306]▼▲	-3.727 0.060▼▼	1	6	19	20	-
$TLI\varepsilon_t$	3.167▲**▲ (0.989)▲▼ [3.483]▼▲	3.076 0.091▼▲	22	4	-	22	-2	3.122▲**▲ (0.971)▲▼ [3.487]▼▲	3.076 0.046▼▲	22	4	-	22	-2
$M\varepsilon_t$	0.664▼ (0.103)▲ [6.769]▼	0.679 0.015▲	26	-	-	26	-	0.664▼ (0.101)▲ [6.977]▼	0.679 0.015▲	26	-	-	26	-
$IM\varepsilon_t$	-							0.187▼** (0.140)▲ [1.761]▼	0.222 0.035**▲	13	12	1	14	-2

Table 10.1: Summary Of Unrestricted Model Results (Continued…)

Panel B: Goodness-Of-Fit And Model Selection Criteria						
	Unrestricted Market Model			Unrestricted Model		
	Mean Value	Minimum	Maximum	Mean Value	Minimum	Maximum
$\bar{R}^2$	0.310▼***▲***	0.111 Fixed line telecom.	0.653 Mining	0.322▼***▲***	0.114 Automobiles & parts	0.678 Mining
AIC	-2.985▲***▼***	-3.809 Food producers	-1.751 Industrial metals &	-3.001▲***▼***	-3.798 Food producers	-1.753 Industrial metals & mining
BIC	-2.789▲***▼***	-3.622 Food producers	-1.548 Industrial metals & mining	-2.789▲***▼***	-3.595 Food producers	-1.532 Industrial metals & mining

Notes:
The asterisks, ***, ** and *, indicate statistical significance at the respective 1%, 5% and 10% levels of significance. All factors are in innovations (unexpected changes) (Section 6.2.2; Section 6.2.3.; Table 6.4.), where BP_{t-1} - Building Plans Passed, $LEAD_{t-1}$ - Leading Indicator, BUS_t - Business Activity, USD_t - Rand-Dollar Ex. Rate, MET_t - Metal Prices, LTY_t - Long-Term Gov. Bond Yields, TLI_t - Trading Partner Lead. Index, R_{Mt} - JSE All Share Index and R_{IMt} - MSCI World Index (US$). In Panel A, Mean Coeff. is the mean value of the intercept and the coefficients associated with each factor. Values in the parentheses ( ) are the mean coefficient standard errors (Std Error) and the values in the brackets [ ] are the mean *z*-scores (|Z-score|). In the third column, Mean LS Co. are the mean values of least squares intercepts and coefficients of the benchmark model. |Diff.| are the absolute mean differences between ML and least squares coefficients. $b_{ik} > 0$ and $b_{ik} < 0$ indicate the respective number of coefficients that are statistically significant and have a positive or negative impact. Total Sig. is the total number of statistically significant coefficients associated with each factor across the return series in the sample. **Δ**Sig is the increase or decrease in the number of statistically significant coefficients relative to the benchmark model specification. In Panel B, Mean is the arithmetic mean of the $\bar{R}^2$, AIC and BIC values across sectors. The Minimum and Maximum values correspond to the lowest and highest values observed and the associated sectors for which they are observed. Across panels, the first ▲ or ▼ symbol indicates that a value is larger or smaller relative to

Finally, intercepts are statistically significant for 16 and 15 sectors for the two unrestricted specifications respectively. For the unrestricted market model, the number of statistically significant intercepts is comparable to that of the restricted specification. For the unrestricted model, the number of statistically significant intercepts is lower relative to that for the restricted specification (-1) but still above that of the benchmark specification (+2). The fewer significant intercepts for the unrestricted specification again implies that $M\varepsilon_t$ by itself fails to account for all omitted influences and that there is a potential role for a second residual market factor, although its contribution is likely to be marginal. These results suggest that the intercepts continue to be biased relative to those of the benchmark model and that the significance of factors continues to be understated.

On an individual basis, for the unrestricted market model, the number of statistically significant coefficients is lower relative to that for the benchmark specification for BP_{t-1} (10 to 7 [-3]), $LEAD_{t-1}$ (17 to 13 [-4]), BUS_t (18 to 15 [-3]), $USD\varepsilon_t$ (16 to 13 [-3]) and $TLI\varepsilon_t$ (24 to 22 [-2]). For the unrestricted model, the number of statistically significant coefficients for certain factors is also lower relative to that for the benchmark specification. Factors for which significance remains understated are BP_{t-1} (10 to 9 [-1]), $LEAD_{t-1}$ (17 to 13 [-4]), BUS_t (18 to 15 [-3]), $USD\varepsilon_t$ (16 to 14 [-2]) and $TLI\varepsilon_t$ (24 to 22 [-2]). Although the general level of discrepancy between the number of statistically significant coefficients for the benchmark specification and the unrestricted specifications is generally lower relative to that observed for the restricted model, inferences relating to the statistical significance of most factors remain impacted.

Encouragingly, the results indicate that while the impact of the abovementioned factors is still understated, the general inference that $LEAD_{t-1}$ and $USD\varepsilon_t$ are systematic in nature is in line with the inferences drawn from results of the benchmark model (Table 8.1). Furthermore, it appears that the inclusion of the second residual market factor in the unrestricted model reduces the discrepancies between the number of significant coefficients for specific factors observed for the benchmark and unrestricted model. This is the case for BP_{t-1}, $LEAD_{t-1}$ and $USD\varepsilon_t$ .Surprisingly, the number of sectors that are significantly related to MET_t increases by one and two respectively for the unrestricted market model and the unrestricted model relative to the number of significant instances observed for the benchmark specification.

In summary, the use of residual market factors appears to partially resolve the associated structural misidentification problem attributable to incorrect inferences associated with underspecification. All systematic factors are now identified as such although there is some understatement in terms of significance for certain factors. The APT relation will no longer be underspecified as factors that are pervasive, and correctly identified as such, will be included in the APT relation (Ferson & Harvey, 1994: 785; Elton *et al.,* 1995: 1239).

10.3.2. Coefficient Magnitude Comparisons

The inclusion of $M\varepsilon_t$ and $IM\varepsilon_t$ in the respective specifications produces results that differ from those of the restricted specification. The mean intercepts for the two unrestricted models are still statistically significantly higher (both 0.007) than those for the benchmark model (0.006) but are significantly lower than those for the restricted specification (0.008). It follows that if the intercepts reflect omitted factors, then the intercepts will be biased. Therefore, while these models may still be underspecified, they appear to be an improvement over the restricted specification (Dominguez, 1992: 94). The residual market factors seemingly account for some omitted factors and this translates into a decrease in the mean intercepts.

The number of factors for which mean coefficients differ significantly from those of the benchmark model under the unrestricted specifications is comparable. There is no longer a statistically significant difference between the mean coefficients for BP_{t-1} for both specifications but the coefficients for $TLI\varepsilon_t$ are significantly higher than those for the benchmark specification (3.167 and 3.122 for the respective unrestricted models and 2.865 for the benchmark model). The coefficients for BUS_t for the respective unrestricted specifications are now significantly lower than those of the benchmark specification (0.070 and 0.072 for the respective unrestricted models and 0.079 for the benchmark model). This is not the case for this factor in the restricted specification. The results are somewhat ambiguous. Whereas the inclusion of $M\varepsilon_t$ in the unrestricted market model results in what appears to be a more accurate approximation of the coefficients for BP_{t-1} relative to the restricted model, the coefficients for $TLI\varepsilon_t$ and now BUS_t differ from those of the benchmark specification. The inclusion of $IM\varepsilon_t$ does not appear to yield a significant improvement although the coefficients for this factor differ significantly in the unrestricted model from those of the benchmark model.

Similarly, the mean coefficients for BP_{t-1}, $LEAD_{t-1}$, and LTY_t obtained from both unrestricted models are significantly higher (0.035 and 0.037, 0.925 and 0.921; -3.658 and -3.666 respectively) than those obtained from the restricted model (0.025; 0.837 and -4.037 (not in absolute terms for LTY_t) respectively). Coefficients are significantly lower for BUS_t (0.070 and 0.072)[147] in both unrestricted specifications relative to the coefficients obtained from the restricted model (0.078). This suggests that the inclusion of $M\varepsilon_t$ and $IM\varepsilon_t$ results in coefficient estimates that more closely approximate the benchmark model coefficients, as coefficients now significantly differ from those of the restricted model but not those of the benchmark model. As before, the inclusion of $IM\varepsilon_t$ does not change the results significantly.

The comparison of the differences between the ML estimates for the unrestricted models and least squares coefficient estimates for the benchmark model produces encouraging results, which favour the partial adequacy of the residual market factors. With the exception of the intercepts for both specifications, for which the differences are statistically significant and higher (differences of 0.001 for both specifications, as indicated by the asterisks) than that for the benchmark model (0.0006), none of the differences in the mean coefficients for the macroeconomic factors in the unrestricted specifications are statistically significant. In other words, mean coefficients associated with the macroeconomic factors in the unrestricted models approximate the least squares benchmark model coefficients, which are hypothesised to be best linear unbiased (BLU), if not efficient (Section 10.5.), estimates.

In the unrestricted model, the difference between ML and least squares coefficients is statistically significant for a single (non-macroeconomic) factor, $IM\varepsilon_t$. This is not the case for $IM\varepsilon_t$ in Panel A of Table 8.1. Moreover, the results of comparisons to the least squares coefficients for both unrestricted specifications are in contrast from those in the third column in Panel A of Table 8.1. for $USD\varepsilon_t$, MET_t and $TLI\varepsilon_t$.[148] The differences between ML and least squares benchmark model coefficient estimates are statistically significant for these factors in Table 8.1. Overall, the unrestricted specifications appear to more closely approximate least squares coefficient estimates which are theoretically unbiased (Lee &

[147] Based upon the results of the Wilcoxon matched-pairs signed-rank test for differences in the medians of coefficients from the restricted and unrestricted specifications.

[148] The results in Table 8.1. are somewhat ambiguous. According to the Wilcoxon matched-pairs signed-ranked test, the difference between ML and least squares benchmark coefficients for $USD\varepsilon_t$ is statistically significant. For MET_t, the results of the Wilcoxon matched-pairs sign-ranked test contradict those of the *t*-test indicating that differences are not statistically significant.

Lemieux, 2010: 286). However, a closer examination suggests that differences between the ML and least squares coefficients for the unrestricted specifications are generally, although not always, larger relative to those for the benchmark specification. Factors for which differences are larger are BP_{t-1} (0.003, 0.003 for the unrestricted models and 0.001 for the benchmark model respectively), $LEAD_{t-1}$ (0.042, 0.038 and 0.022 respectively), BUS_t (0.009, 0.007 and 0.0006 respectively), $USD\varepsilon_t$ (0.010, 0.007 and 0.006 respectively) , $M\varepsilon_t$ (0.015, 0.015 and 0.015 respectively (marginally)) and $IM\varepsilon_t$ (0.035 for the unrestricted model and 0.005 for the benchmark model). In contrast, the differences between ML and least squares coefficients for the unrestricted specifications are smaller relative to those for the benchmark specification for MET_t (0.013, 0.015 and 0.018 respectively), LTY_t (0.069, 0.060 and 0.193 respectively) and $TLI\varepsilon_t$ (0.091, 0.049 and 0.211 respectively). As evident from the above, the inclusion of the second residual market factor modestly reduces the differences between ML and least squares coefficient estimates in most instances.

The inclusion of the residual market factors in both specifications also reduces differences relative to those observed for the restricted specification. Relative to the restricted model, differences are lower for the unrestricted models for the intercepts (0.001, 0.001 for the respective unrestricted models and 0.002 for the restricted model), BP_{t-1} (0.003, 0.003 and 0.014 respectively), $LEAD_{t-1}$ (0.042, 0.038 and 0.046 respectively), $USD\varepsilon_t$ (0.007 for the unrestricted model (marginally) and 0.007 for the benchmark model), MET_t (0.013, 0.015 and 0.016 respectively) and LTY_t (0.069, 0.060 and 0.311 respectively). Differences are greater for BUS_t (0.009, 0.007 and 0.0004 respectively) and $TLI\varepsilon_t$ (0.091, 0.046 and 0.026 respectively).

The present discussion indicates that, overall, the inclusion of the residual market factors reduces coefficient bias relative to the restricted specification. A possible explanation is that the residual market factors reflect previously omitted information and this impacts the structure of the conditional variance which enters coefficient estimates (Section 10.5.). In the unrestricted market model, most coefficients on the macroeconomic factors do not differ significantly from those of the benchmark model. The same may be said about the unrestricted model. However, coefficients differ significantly from those of the restricted model for four of the macroeconomic factors in the unrestricted market model. This suggests that the estimated coefficients in the unrestricted market model more closely approximate those of benchmark model than those of the restricted model. Overall, the benchmark model produces lower differences between the ML and least squares coefficient estimates for most factors relative to the unrestricted market model and the unrestricted model. However, differences are not significant, unlike those for the benchmark model for which differences are statistically significant for a single factor and ambiguous for an additional two factors. In general, the differences between ML and least squares coefficients derived from the unrestricted models are smaller relative to those of the restricted model. This suggests that the inclusion of the residual market factors reduces bias. The greatest contribution comes from the inclusion of $M\varepsilon_t$ and the contribution of $IM\varepsilon_t$ is marginal.

10.3.3. Model Assessment And Comparisons

It is expected that the mean $\bar{R}^2$ s for both unrestricted specifications will be higher than that of the restricted specification. By accounting for omitted factors, the residual market factors will contribute to explaining the systematic variation in the returns. What is of particular interest is whether the respective $\bar{R}^2$ s of the unrestricted specifications approximate the mean $\bar{R}^2$ of the benchmark specification.

The mean $\bar{R}^2$ for the unrestricted market model is 0.310 and 0.322 for the unrestricted specification. The mean $\bar{R}^2$ for both specifications is more than double that of the restricted specification (mean $\bar{R}^2$ = 0.142). However, both $\bar{R}^2$ s are below 0.504, the mean $\bar{R}^2$ of the benchmark model, suggesting that both specifications fail to approximate the explanatory power of the benchmark specification. Significance tests confirm that the differences between the mean $\bar{R}^2$ s for the unrestricted specifications and the benchmark and restricted models are statistically significant. This indicates that although the inclusion of the residual market factors in the unrestricted models improves explanatory power, the explanatory power of these two specifications still fails to approximate that of the benchmark model. Furthermore, the increase in the mean $\bar{R}^2$ from 0.310 for the unrestricted market model to 0.322 for the unrestricted model suggests that most of the explanatory power associated with omitted factors is reflected in the conventional residual market factor, $M\varepsilon_t$. The contribution of $IM\varepsilon_t$ is marginal, if not negligible.

The $\bar{R}^2$ for the unrestricted market model ranges between 0.653 for the mining sector and 0.111 for the fixed line telecommunications sector. For the unrestricted model, the $\bar{R}^2$

ranges between 0.678 for the mining sector and 0.114 for the automobiles and parts industrial sector. The range of the $\bar{R}^2$ values for both specifications is similar, again suggesting that the contribution of the second residual market factor is marginal. In comparison, the $\bar{R}^2$ for the benchmark model ranges between 0.171 for the fixed line telecommunications sector and 0.941 for the mining sector. For the restricted specification, the $\bar{R}^2$ values range from 0.246 for the general retailers sector to 0.032 for the fixed line telecommunications sector. The respective ranges of the $\bar{R}^2$ values for the unrestricted models shift upwards relative to the restricted model but do not approximate that of the benchmark model.

The mean AIC value for the unrestricted market model is -2.985 and ranges between -1.751 for the industrial metals and mining sector and -3.809 for the food producers sector. The mean AIC value for the unrestricted model is only marginally lower at -3.001 and ranges between -1.753 for the industrial metals and mining sector and -3.798 for the food producers sector. The mean BIC value for the unrestricted market model is -2.789 and values range between -3.622 for the food producers sector and -1.548 for the industrial metals and mining sector. The mean BIC value for the unrestricted model is -2.789 and ranges between -3.595 for the food producers sector and -1.532 for the industrial metals and mining sector. The unrestricted specifications significantly underperform the benchmark specification. The mean AIC value for the benchmark model is -3.348 and AIC values range between -4.956 for the mining sector and -2.035 for the fixed line telecommunications sector. The mean BIC value is -3.114 and the BIC values range between -4.736 for the mining sector and -1.797 for the fixed line telecommunications sector. While the inclusion of the residual market factors results in an improvement in terms of the ability of the specifications to replicate the return series and yields a closer approximation of the true return generating process, both specifications significantly underperform the benchmark specification (Mills & Markellos, 2008: 34; Spiegelhalter *et al.,* 2014: 1). Additionally, the closely comparable mean AIC and BIC values and the associated ranges for both unrestricted specifications again suggest that the inclusion of $IM\varepsilon_t$ has a marginal impact on improving predictive accuracy and the approximation of the true return generating process. The inclusion of the residual market factors in the unrestricted specifications however yields significant improvements relative to the restricted specification, as evident from the significantly lower mean AIC and BIC statistics in Panel B of Table 10.1. relative to those in Panel B of Table 9.1. for the restricted model.

The inclusion of $M\varepsilon_t$ in the unrestricted market model yields the largest improvements in the overall $\bar{R}^2$, AIC and BIC values. Nevertheless, the unrestricted market model underperforms the benchmark model according to these measures. This suggests that $M\varepsilon_t$ is not an adequate proxy for omitted factors. A specification that incorporates only macroeconomic factors and a conventional residual market factor will continue to exhibit symptoms of underspecification. This is not resolved by the inclusion of $IM\varepsilon_t$ and suggests that the contribution of a second residual market factor is marginal – although extant.

10.4. MODEL DIAGNOSTICS AND ROBUSTNESS

Table 10.2. reports the abridged results of the diagnostic tests for the unrestricted specifications. *F*-statistics are statistically significant for all industrial sectors, for both unrestricted specifications. This confirms the significance of combining the macroeconomic factors with the residual market factor in the unrestricted market model and combining the macroeconomic factors and the two residual market factors in the unrestricted model (Sadorsky & Henriques, 2001: 204). The resultant mean *F*-statistics are lower than that of the benchmark specification, with the mean *F*-statistic for the unrestricted market model equalling 16.143 and for the unrestricted model equalling 15.530.

Table 10.2: Abridged Unrestricted Model Diagnostics

Test	Mean Value	Total Sig.	ΔSig	Mean Value	Total Sig.	ΔSig
	Unrestricted market model			**Unrestricted model**		
F-Test	16.143▼▲	26/26	-	15.530▼▲	26/26	-
JB Test	17.862▲▲	12/26▼▼	-3	15.229▲▼	14/26▼-	-1
$Q(1)$	1.365▼▼	4/26▼▼	-1	1.322▼▼	3/26▼▼	-2
$Q(5)$	6.412▼▼	3/26▼▼	-1	6.614▼▼	5/26▲▼	+1
$Q^2(1)$	0.478▲▼	0/26	-	0.406▲▼	0/26	-
$Q^2(5)$	3.517▼▼	0/26	-	3.686▼▲	0/26	-
ARCH(1)	0.469▲▲	0/26	-	0.398▲▼	0/26	-
ARCH(5)	0.653▼▼	0/26	-	0.712▼▼	0/26	-

Notes:
Significance is recorded at the 10% level of significance. *F*-Test reports the results for Wald's test of linear restrictions jointly equating all explanatory factors in the respective specifications to zero. JB Test summarises the results of the Jarque-Bera test for normality. $Q(1)$ and $Q(5)$ are Ljung-Box *Q*-statistics indicating whether serial correlation in the residuals is statistically significant at the first order and jointly up to five orders of serial correlation respectively. $Q^2(1)$ and $Q^2(5)$ are Ljung-Box test statistics for non-linear dependence in the residuals at the first order and jointly up to five orders. ARCH(1) and ARCH(5) are Lagrange Multiplier (LM) tests for ARCH effects in the residuals at the first and fifth orders respectively. Mean Value reports the mean of the respective test statistics and Total Sig. reports the number of instances in which the results of the respective tests applied are statistically significant. **Δ**Sig is the increase or decrease in the number of statistically significant coefficients relative to the benchmark model specification. The first ▲ or ▼ symbol indicates that a value is larger or smaller relative to that observed for the benchmark model. The second ▲ or ▼ symbol indicates that a value is larger or smaller that observed for the restricted model.

This is an improvement over the restricted specification, which has a mean *F*-statistic of 6.692 but is still far below that of the benchmark model, which has a mean *F*-statistic of 36.528. As *F*-statistics are estimated from differences between the sums of squared residuals (equation 6.30), the increase in the mean *F*-statistics implies that the sum of squared residuals of the unrestricted specifications is lower relative to that of the restricted specification but greater relative to that of the benchmark model (Blackwell, 2008: 4). This is expected; the inclusion of the residual market factors, by accounting for omitted factors, decreases the sum of squared residuals which will reflect omitted factors. The inclusion of the factor scores in the benchmark model further decreases the sum of squared residuals. The inclusion of the second residual market factor, $IM\varepsilon_t$, does not improve overall significance, as suggested by the decline in the mean *F*-statistic associated with the unrestricted model relative to the unrestricted market model.

In summary, the results for individual sectors confirm the multifactor structure of a return generating process that combines first a single residual market factor and then a second residual market factor, together with the macroeconomic factors. These results also suggest that the inclusion of a residual market factor or even two residual market factors fails to achieve the same levels of significance as the benchmark model (see Sullivan & Feinn, 2012: 279 for a discussion of effect size that can be applied to these results).

The residual market factors do not appear to have a clear impact on the conditional normality of the residuals for individual sectors and a clear pattern across individual sectors does not emerge. The residual series of 12 and 14 industrial sectors for the respective unrestricted models in Table 10.2. exhibit significant departures from normality. For the restricted specification, the residuals of 14 sectors are associated with significant departures from normality and for the benchmark specification, 15 residual series exhibit departures from normality. This is compounded by a finding that for certain sectors, residual series for the unrestricted market specification that are now conditionally normal are not conditionally normal for the restricted specification and *vice versa*. Examples of sectors that now exhibit departures from normality are the electronic and electrical equipment, fixed line telecommunications and equity and investment instruments sectors. The sectors for which residuals are no longer conditionally normal are the pharmaceuticals and biotechnology and life insurance sectors. For the unrestricted specification and as an example, the residual series that are now conditionally normal are for the electronic and electrical equipment sector and the equity investment and instruments sector. The residuals for the life insurance and software and computer services sectors now depart from conditionally normality (see Table A1.2. and Table A1.3. in Appendix A).

In contrast, an interpretation of the mean JB test statistics that take into account effect size (given that the sample size does not change, see Sullivan & Feinn, 2012: 279) provides some evidence that factor omission impacts the residual distribution. The mean JB test statistics for both unrestricted specifications, 17.862 for the unrestricted market model and 15.229 for the unrestricted model, are larger than that of the benchmark model of 11.511. This suggests that the inclusion of the factor analytic set in the benchmark model impacts the skewness and kurtosis coefficients that determine the individual JB test statistics (equation (6.4)), and the subsequent (potential) rejection of the null hypothesis of a normally distributed series. However, a comparison of these mean JB test statistics to that of the restricted specification (15.868) challenges the notion that factor omission impacts the residual distribution. The mean JB test statistic for the unrestricted market model is higher than that of the restricted market model and the mean JB test statistic for the unrestricted specification is comparable to that of the restricted model. This is contrary to expectations if residual non-normality is attributable to omitted factors (see Downing & Clark, 2010: 403).

These findings suggest that $M\varepsilon_t$ fails to account for outliers that may be associated with omitted factors, which are otherwise explained by the factor analytic augmentation in the benchmark specification. The inclusion of $IM\varepsilon_t$ results in a marginally lower mean JB test statistic. As the mean JB test statistics for the unrestricted specifications are either slightly higher than that of the restricted specification or comparable, it remains difficult to conclude that departures from normality are fully attributable to factor omission. These results, although suggestive, are ambiguous.

The inclusion of $M\varepsilon_t$ appears to address the residual serial correlation that is observed in the residuals of the restricted specification. Encouragingly, the number of instances of statistically significant residual serial correlation is only marginally higher than of the benchmark model (Table 8.2.). The $Q(1)$ and $Q(5)$ statistics indicate that in contrast to the restricted specification for which the residuals of 12 industrial sectors exhibit significant residual serial correlation at both the lower and/or higher orders, seven sectors exhibit evidence of residual serial correlation for both unrestricted specifications. This is slightly higher than the six sectors for the benchmark model (see Appendix A). The inclusion of $IM\varepsilon_t$ does not seem to reduce instances of significant residual serial correlation further as seven sectors continue to exhibit significant serial residual correlation for both unrestricted specifications. These results imply that the higher number of instances of significant serial correlation observed for the restricted model at lower and/or higher orders is attributable to impure serial correlation associated with omitted factors (Mutsune, 2008: 6; Studenmund, 2014: 325).

No clear pattern emerges for the sectors that still exhibit significant residual serial correlation. Only three sectors exhibit significantly serially correlated residuals across both unrestricted specifications. These are the industrial metals and mining, fixed line and telecommunications and banks sectors. When the number of individual significant $Q(1)$ and $Q(5)$ statistics is considered for the unrestricted market model, the number of significant instances is marginally lower relative to that for the benchmark model (-1 for $Q(1)$ and $Q(5)$ respectively). For the unrestricted model, two fewer $Q(1)$ statistics are significant relative to the benchmark model and for the $Q(5)$ statistics, a single additional series exhibits significant serial residual correlation up to the fifth order of serial correlation.

The magnitude of the mean Q-statistics suggests that the inclusion of $M\varepsilon_t$ and $IM\varepsilon_t$ in the unrestricted market model and the unrestricted model translates into a reduction in joint serial correlation relative to the restricted model. The respective mean $Q(1)$ statistics for these specifications are 1.365 and 1.322 and the mean $Q(5)$ statistics are 6.412 and 6.614, respectively. For the restricted specification, the respective mean $Q(1)$ and $Q(5)$ statistics are 1.919 and 7.107. These results suggest that the inclusion of $M\varepsilon_t$ reduces impure serial correlation. The contribution of $IM\varepsilon_t$ seems to be negligible. In summary, the inclusion of $M\varepsilon_t$ in the macroeconomic linear factor model should sufficiently address potentially misleading inferences arising as a result of residual serial correlation associated with omitted factors. Finally, and interestingly, for the benchmark model, the mean $Q(1)$ statistic is marginally higher at 1.744 than those of the unrestricted models but marginally lower at 6.585 relative to the mean $Q(5)$ statistics for these specifications. This result is ambiguous.

Even though the serial correlation in the residuals of a number of industrial sectors may be pure in nature and is not resolved by the inclusion of the residual market factors, it may affect standard errors, *z*-scores and resultant *p*-values (Brauer & Gómez-Sorzano, 2004: 38). To confirm the robustness of these results, the unrestricted specifications for which the residual series are serially correlated at either of the orders tested are re-estimated with Newey-West HAC standard errors.[149] For the unrestricted market model, three sectors are impacted. These are the forestry and paper industrial sector for which LTY_t is now statistically significant, the industrial metals and mining sector for which LTY_t and $TLI\varepsilon_t$ are no longer statistically significant and the banks industrial sector for which $TLI\varepsilon_t$ is no longer significant. The impact on the results of the unrestricted specification is more extensive; five sectors are impacted. For the industrial metals and mining sector, LTY_t , $TLI\varepsilon_t$ and $IM\varepsilon_t$ are no longer statistically significant and for the construction materials sector, $TLI\varepsilon_t$ is significant but $IM\varepsilon_t$ is no longer significant. For the health care equipment and services sector, $LEAD_{t-1}$ and BUS_t are no longer statistically significant. For the banks sector, $TLI\varepsilon_t$ is now statistically significant whereas for the general financial sector, BP_{t-1} is no longer significant but BUS_t is now statistically significant. Even though these discrepancies exist

[149] Results are available upon request.

and are potentially attributable to the impact of residual serial correlation, they are limited and non-systematic and therefore do not affect the main findings of this chapter. This is especially applicable to LTY_t and $TLI\varepsilon_t$, which are widely significant across the unrestricted specifications.

In line with expectations and following the application of the appropriate ARCH(p) and GARCH(p,q) models, the $Q^2(1)$and $Q^2(5)$ statistics indicate that the residual series are free of non-linear serial correlation. ARCH(1) and ARCH(5) LM tests confirm that the ARCH(p) and GARCH(p,q) specifications ensure that the residuals do not reflect ARCH effects. As with the results in Chapter 9, the mean $Q^2(1)$and $Q^2(5)$ statistics are suggestive, but not revealing. The respective mean $Q^2(1)$ statistics for the unrestricted models (0.478 and 0.406) are lower than that of the restricted model (0.524) but higher than that of the benchmark model (0.236). However, the respective mean $Q^2(5)$ statistics for the unrestricted models (3.517 and 3.686) are comparable to that of the restricted specification (3.638). In contrast, the mean $Q^2(5)$ for the benchmark specification (3.812) is somewhat higher than the mean $Q^2(5)$ statistics for both models. This suggests that a rudimentary effect size analysis makes it difficult to pronounce on how the inclusion of the residual market factors impacts the residual variance structures of the series in the sample. This question is addressed in detail in Section 10.5.

The results of this section show that the mean F-statistics are lower than that of the benchmark model but greater than that that of the restricted specification. This implies that there is an effect size associated with incorporating $M\varepsilon_t$ and then $IM\varepsilon_t$ into the unrestricted models . The same holds for the factor analytic augmentation in the benchmark model. Significant F-statistics for all sectors confirm that combining the macroeconomic factors with $M\varepsilon_t$ and subsequently $IM\varepsilon_t$ translates into a valid multifactor representation of the return generating process. The number of significant deviations from normality for the residuals of the unrestricted specifications does not differ substantially from those observed for the restricted and benchmark specifications. The impact of $M\varepsilon_t$ and $IM\varepsilon_t$ on the residual distribution is inconclusive. The inclusion of $M\varepsilon_t$ in the linear factor model reduces residual serial correlation overall and this is evidence favouring the efficacy of the conventional residual market factor. The number of instances of statistically significant residual serial correlation is comparable to that of the benchmark specification. The contribution of $IM\varepsilon_t$ remains minor. As before, the residuals are free of non-linear dependence and do not exhibit ARCH effects.

10.5. VARIANCE AND CONDITIONAL VARIANCE

The increase in the number of statistically significant coefficients relative to that for the restricted model suggests that the inclusion of the residual market factors translates into efficiency gains (Section 10.3.1.). Although the mean standard errors in Panel A of Table 10.1. for both unrestricted models are still larger than those of the benchmark specification, they are lower in magnitude relative to those of the restricted specification. For example, the mean standard errors for BP_{t-1} increase from 0.031 for the benchmark model to 0.042 for the restricted model and then decrease to 0.037 for both unrestricted specifications, the unrestricted market model and the unrestricted model, respectively. Similarly, the mean standard errors for $LEAD_{t-1}$ are 0.425 for the benchmark model and 0.488 and 0.486 for the respective unrestricted specifications but 0.537 for the restricted specification. These observations apply to all factors; mean standard errors are greater for each factor for both unrestricted specifications relative to the benchmark model but lower relative to the restricted model. The inclusion of $IM\varepsilon_t$ in the unrestricted model has an almost negligible impact on mean standard errors for some factors. For example, the mean standard error for $USD\varepsilon_t$ decreases from 0.156 in the restricted model to 0.131 in the unrestricted market model and to 0.130 in the unrestricted model. For BUS_t, the mean standard error decreases from 0.049 for the restricted model to 0.045 with the inclusion of $M\varepsilon_t$, and to 0.044 with the inclusion of $IM\varepsilon_t$. With the exception of LTY_t, for which the mean standard errors increase in unrestricted model, the efficiency gains from including a second residual market factor appear to be negligible. [150]

Mean *z*-scores increase for all factors in the unrestricted specifications relative to the restricted specification. For example, *z*-scores for BP_{t-1} increase in the unrestricted market model and the unrestricted model relative to the restricted model from 0.922 to 1.147

[150] In Panel A of Table 8.1, the significance of $IM\varepsilon_t$ is limited to 16 sectors and for the unrestricted model in Table 10.3, it is limited to 14 sectors. This suggests that the inclusion of $IM\varepsilon_t$ may be associated with overstated standard errors in the unrestricted model for some coefficients as a result of redundant factor bias (Studenmund , 2014: 186).

(24.403% increase) and to 1.187 (26.741%) respectively. For $LEAD_{t-1}$, z-scores increase from 1.609 to 2.026 (25.917% increase) and 2.038 (26.663%) respectively. For MET_t, z-scores increase from 2.934 to 3.344 (13.974%) and 3.306 (12.679%) respectively. That increases in z-scores are driven by decreases in standard errors as opposed to pure increases in coefficient magnitudes is confirmed by increases in the z-scores for factors which show a decrease in mean coefficients for both unrestricted models. For example, the mean z-scores for BUS_t increase from 1.667 to 1.782 (6.899%) and 1.824 (9.418%) respectively. Yet, for the unrestricted market model, the mean coefficient for BUS_t decreases from 0.078 for the restricted model to 0.070 (10.256% decrease) for the unrestricted market model and 0.072 (8.333%) for the unrestricted model. For $USD\varepsilon_t$, z-scores increase from 1.422 for the restricted model to 1.683 (18.354%) and 1.743 (22.574%) respectively. The mean coefficient for $USD\varepsilon_t$ decreases (in absolute terms) from 0.167 in the restricted model to 0.163 (2.395% decrease) in the unrestricted market model. The same can be said for LTY_t , for which the mean coefficient decreases (in absolute terms) from 4.037 to 3.658 (9.388%) and 3.666 (9.190%) for the respective unrestricted models but z-scores increase from 2.934 to 3.344 (13.974%) and 3.306 (12.679%) respectively. These results suggest that increases in z-scores for both specifications are mostly driven by decreases in the size of the standard errors relative to those of the restricted specification (as indicated by the second ▼ reported next to the mean standard errors for each factor in brackets in Table 10.1.) as opposed to sole increases in coefficient magnitudes.

Increases in z-scores are also partly driven by increases in the magnitude of estimated coefficients. For example, for both unrestricted specifications, the mean coefficients for BP_{t-1} , $LEAD_{t-1}$, MET_t and $TLI\varepsilon_t$ are larger relative to the mean coefficients for these factors in the restricted specification (as indicated by the second ▲ symbol in the second column of Table 10.1.). Increases in coefficient magnitudes re-enforce the impact of smaller standard errors on z-scores. A noteworthy finding in the context of this study is that increases in z-scores are also driven by changes in coefficient magnitudes . This suggests that incorporating the characteristics of heteroscedasticity into the estimation of linear factor models impacts model coefficient estimates and their significance. Within the ARCH/GARCH framework, underspecification has a dual impact, namely on residual variance and on coefficient estimates (Hamilton, 2010; Bucevska, 2011: 631). By including

$M\varepsilon_t$ in the unrestricted market model and $IM\varepsilon_t$ in the unrestricted model, the structure of heteroscedasticity will be impacted. If changes in the structure of heteroscedasticity are responsible for increases in coefficient magnitudes, then it can be concluded that this also contributes to increasing *z*-scores (Section 6.4.2.).

The present discussion suggests that the inclusion of $M\varepsilon_t$ translates into efficiency gains relative to the restricted model although mean standard errors for the unrestricted market model are still higher relative to those for factors in the benchmark specification. This can be attributed to an upward bias in residual variance. Most of the increases in the *z*-scores are attributable to the incorporation of $M\varepsilon_t$ in the unrestricted market model. The overall increase[151] in the mean *z*-scores across the macroeconomic factors, relative to the restricted model following the inclusion of $M\varepsilon_t$ in the unrestricted market model, is 18.669%. When $IM\varepsilon_t$ is incorporated in the unrestricted model, the overall increase is 20.545% relative to the restricted model. The difference between these increases is 1.876%, a testimony to the marginal contribution of $IM\varepsilon_t$.

[151] Defined as the mean of individual percentage increases across the mean *z*-scores.

Table 10.3: Unrestricted Model Residual Variance And Variance Structure

Panel A: Residual Variance

	Unrestricted Market Model			**Unrestricted Model**		
	Mean Value	**Minimum**	**Maximum**	**Mean Value**	**Minimum**	**Maximum**
$\sigma^2_{MR/UR\varepsilon_i}$	0.003484▲***▼***	0.001165 Food producers	0.012903 Industrial metals & mining	0.003411▲***▼***	0.001164 Food producers	0.012748 Industrial metals & mining
Sig. $\sigma^2_{MR/UR\varepsilon_i} > \sigma^2_{BM\varepsilon_i}$	18/26▼			13/26▼		
$\sigma^2_{MR/UR\varepsilon_i} = \sigma^2_{BM\varepsilon_i}$	8/26▲			13/26▲		
$\sigma^2_{MR/UR\varepsilon_i} = \sigma^2_{RES\varepsilon_i}$	19/26			15/26		
Sig. $\sigma^2_{MR/UR\varepsilon_i} < \sigma^2_{RES\varepsilon_i}$	7/26			11/26		

Panel B: Conditional Variance Structure

	Unrestricted Market Model			**Unrestricted Model**		
Model	**ARCH(1)**	**GARCH(1,1)**	**GARCH(2,1)**	**ARCH(1)**	**GARCH(1,1)**	**GARCH(2,1)**
Frequency	**13**	**12**	**1**	**15**	**10**	**1**
	Mean Coeff.	**Mean Coeff.**	**Mean Coeff.**	**Mean Coeff.**	**Mean Coeff.**	**Mean Coeff.**
ω	0.003 (13)	0.0003 (2)	0.001	0.002 (15)	0.0003 (2)	0.001
α_1	0.135 (6)	0.107 (4)	-	0.132 (6)	0.115 (4)	0.262 (1)
α_2			0.291 (1)			
β_1		0.809 (12)	0.499 (1)		0.806 (10)	0.507 (1)
Sig F-Test	**6/13**	**12/12**	**1/1**	**6/15**	**10/10**	**1/1**

Table 10.3: Unrestricted Model Residual Variance And Variance Structure (Continued…)

Panel C: Conditional Heteroscedasticity		
	Unrestricted Market Model	**Unrestricted Model**
α_i	0.123▲▲	0.125▲▲

Notes:
The asterisks, ***, ** and *, indicate statistical significance at the respective 1%, 5% and 10% levels of significance. In Panel A, the Mean Value is the mean of the residual variance across sectors. The Minimum and Maximum are the lowest and highest residual variances observed for the respective sectors. Sig. $\sigma^2_{MR/UR_i} > \sigma^2_{BM\varepsilon_i}$, $\sigma^2_{MR/UR\varepsilon_i} = \sigma^2_{BM\varepsilon_i}$ and $\sigma^2_{MR/UR\varepsilon_i} = \sigma^2_{RES\varepsilon_i}$ $\sigma^2_{MR/UR\varepsilon_i} < \sigma^2_{RES\varepsilon_i}$ are comparisons of the residual variance derived from the restricted model and the benchmark model. The number of significant instances as established by the Brown-Forsythe test is reported as is the number of instances in which no significant differences are observed. The number of significant instances is reported, as is the number of instances in which no significant differences are observed. In Panel B, Frequency is indicative of the number of instances of each ARCH(*p*) or GARCH(*p,q*) model applied. Negative ARCH coefficients, which are indicative of the absence of conditional heteroscedasticity and are therefore statistically insignificant, are rounded to zero in aggregation. This is consistent with the non-negativity constraint ($\alpha_i \geq 0$) placed upon the ARCH coefficient in the ARCH(*p*) specification (see Poon, 2005: 38). The numbers in brackets ( ) next to each mean value indicate the number of statistically significant coefficients for each ARCH(*p*) or GARCH(*p,q*) specification at the 10% level of significance. Sig. *F*-Test reports the number of significant instances of Wald's test of linear restrictions for the ARCH and GARCH coefficients. The null hypothesis is that ARCH and GARCH coefficients are jointly equal to zero. In Panel C, α_i is the arithmetic mean of the ARCH coefficients across sectors. Across Panels, the first ▲ or ▼ symbol indicates that a value is larger or smaller relative to that observed for the benchmark model. The second ▲ or ▼ symbol indicates that a value is larger or smaller relative to that observed for the restricted model. Accompanying asterisks, if present, indicate that differences are statistically significant. A superscript "W" indicates a discrepancy between the results of the *t*-test and the Wilcoxon test.

Encouragingly, the mean residual variance estimates for the unrestricted market model, $\sigma^2_{MR\varepsilon_i}$, and the unrestricted model, $\sigma^2_{UR\varepsilon_i}$, show statistically significant reductions in magnitude relative to those of the restricted specification (as indicated by the second ▼ in Panel A of Table 10.3.). The mean residual variance decreases from 0.004307 for the restricted specification to 0.003484 for the unrestricted market model and 0.003411 for the unrestricted model. For the unrestricted market model, the sector with the lowest residual variance is the food producers sector whereas the sector with the largest residual variance is the industrial metals and mining sector. Both of these sectors also have the respective lowest and highest residual variance in the restricted specification (Table 9.3.). Both minimum and maximum values show a downward shift with the inclusion of $M\varepsilon_t$ relative to the restricted model. The residual variance for food producers decreases from 0.001469 to 0.001165 and from 0.014494 to 0.012903 for the industrial metals and mining sector with the inclusion of $M\varepsilon_t$. The same sectors are also associated with minimum and maximum residual variance values in the unrestricted model. The decrease in residual variance is almost negligible when $IM\varepsilon_t$ is included in the unrestricted specification. The residual variance for the food producers sector is 0.001164 and for the industrial metals and mining sector, the residual variance is 0.012748.

Decreases in residual variance are expected; the inclusion of $M\varepsilon_t$ in the unrestricted market model and the subsequent inclusion of $IM\varepsilon_t$ in the unrestricted model, if not redundant, will reduce the portion of residual variance that is associated with omitted systematic factors. This, in turn, will reduce the upward bias in the standard errors and partially alleviate the erroneous tendency not to reject the null hypothesis (Lehmann, 1990: 72; Wooldridge, 2013: 56, 312). This is suggested by the lower number of deviations in the number of statistically significant instances of factor coefficients for these two specifications relative to the restricted model (see the **Δ**Sig column in Panel A of Table 9.1. and Table 10.1.). Nevertheless, the means of the residual variance estimates for the two unrestricted models, $\sigma^2_{MR\varepsilon_i}$ and $\sigma^2_{UR\varepsilon_i}$, are still significantly higher (as indicated by the first ▲ symbol) than that of the benchmark model, which is 0.002483 (Panel A of Table 8.3.). This suggests that the two residual market factors fail to account for all omitted pervasive factors and this is reflected by an upward bias in residual variance. This also explains the remaining discrepancies between the number of statistically significant factors for the unrestricted specifications and the benchmark specification, noted in Section 10.3.1.

These results show that the use of a single residual market factor, and even a second residual market factor, will not fully mitigate erroneous failures to reject the null hypothesis of no impact on returns and the possibility of rejecting the APT relation if idiosyncratic factors that reflect the influence of omitted systematic factors are used in tests of the APT relation (Dominguez, 1992: 97-98; Huang, Liu, Rhee & Zhang, 2009: 153). This is because the residual variance estimates still appear to be inflated and do not approximate those of the benchmark model.

The Brown-Forsythe test indicates that the inclusion of $M\varepsilon_t$ in the unrestricted market model has an impact on the magnitude of individual residual variance estimates. The residual variance for eight industrial sectors (as opposed to four for the restricted specification, see Section 9.5.) is now not significantly different from that of the benchmark specification ($\sigma^2_{MR\varepsilon_i} = \sigma^2_{BM\varepsilon_i}$) (Panel A of Table 10.3.). The four additional sectors for which residual variance is now comparable are the forestry and paper, beverages, non-life insurance, equity investment and instruments sectors. However, the residual variance for the remaining 18 out of 26 sectors is significantly higher than that of the benchmark specification ($\sigma^2_{MR\varepsilon_i} > \sigma^2_{BM\varepsilon_i}$). This is a minor improvement over the restricted model for which 22 of the 26 sectors in the sample exhibit significantly higher residual variance than that of the corresponding sectors in the benchmark model. The residual variance of the unrestricted market model is significantly lower relative to that of the restricted model for only a limited number of sectors - seven of the 26 industrial sectors ($\sigma^2_{MR\varepsilon_i} < \sigma^2_{RES\varepsilon_i}$). These are the mining, general industrials, beverages, banks, life insurance, general financial and equity investments and instrument sectors (see Table A1.3. in Appendix A). This comparison yields concerning results. The residual variance of 19 sectors is comparable to that of the restricted specification ($\sigma^2_{MR\varepsilon_i} = \sigma^2_{RES\varepsilon_i}$).

The inclusion of $IM\varepsilon_t$ in the unrestricted model yields somewhat of an improvement; 13 sectors, as opposed to 18 sectors for the unrestricted market model, exhibit significantly higher residual variance relative to that of the benchmark model ($\sigma^2_{UR\varepsilon_i} > \sigma^2_{BM\varepsilon_i}$). This means that the residual variance for 13 out of 26 sectors, as opposed to eight sectors for the unrestricted market model, is now comparable to that of the benchmark model ($\sigma^2_{UR\varepsilon_i} = \sigma^2_{BM\varepsilon_i}$) (Panel A of Table 10.3.). The additional sectors (to that of the unrestricted market model) for which residual variance is now comparable to that of the benchmark model are the chemicals, forestry and paper, industrial engineering, media, travel and leisure, and the life insurance sectors (see Table A1.4. of in Appendix A). Somewhat encouragingly, the residual variance of 11 sectors is significantly lower than that of the restricted specification ($\sigma^2_{UR\varepsilon_i} < \sigma^2_{RES\varepsilon_i}$). This is in comparison to seven sectors for which the residual variance is lower relative to the residual variance in the unrestricted market model. The residual variance for 15 sectors is comparable to the residual variance of the restricted model ($\sigma^2_{MR/UR\varepsilon_i} = \sigma^2_{RES\varepsilon_i}$).

The results of the Brown-Forsythe test suggest that the use of the residual market factors reduces upward bias in the residual variance attributable to omitted factors but does not eliminate it. The results of this test also show that the inclusion of $IM\varepsilon_t$ contributes to reducing residual variance and this contribution is over and above that of $M\varepsilon_t$. Nevertheless, the residual variance for most sectors still exhibits an upward bias relative to the benchmark specification for both specifications. This suggests that the residual market factors do not fully account for omitted factors and that the power of statistical tests will continue to be adversely affected. These findings go some way to explain as to why systematic factors may now be correctly identified as such in Section 10.3.1. and why there are still some deviations in the number of significant coefficients for the unrestricted and the benchmark models.

The structure of the conditional variance of the unrestricted models, as described by the ARCH(*p*) and GARCH(*p,q)* models, is summarised in Panel B of Table 10.3. The conditional variance structures of the residuals of the unrestricted market specification in Panel B more closely resemble those of the restricted model in Chapter 9 than those of the benchmark specification in Chapter 8. A total of 13 series are described by the short-memory ARCH(1) model with *F*-statistics confirming overall significance for six sectors. The long-memory GARCH(1,1) model is applied to 12 sectors and the *F*-statistics confirm overall significance for all sectors. Also, a GARCH(2,1) specification is fitted to a single sector, the banks industrial sector and the *F*-test indicates overall significance. This shows that a total of 19 sectors exhibit residual variance that is non-stationary. These results, however, point towards a shift to short-memory conditional variance structures although half of the series are still characterised by long-memory GARCH(*p,q*) specifications. The restricted specification favours long-memory structures; 15 sectors are characterised by the long-memory GARCH(1,1) model. The benchmark model, however, favours short-memory structures; 18 sectors are characterised by the short-memory ARCH(1) model and eight sectors are described by the GARCH(1,1) specifications.

Interestingly, the conditional variance structures underlying the unrestricted model more closely approximate the mostly short-memory structures of the benchmark specification. The ARCH(1) process characterises the conditional variance of 15 sectors, with *F*-statistics confirming overall significance for six out of the 15 sectors. The GARCH(1,1) model describes the conditional variance of 10 sectors and *F*-statistics indicate that this specification is significant for all sectors. A GARCH(2,1) model is also fitted to the banks industrial sector and is significant overall. This suggests that a total of 17 sectors exhibit residual variance that is non-stationary. The prevalence of ARCH(1) models used to describe the conditional variance underlying the unrestricted model points towards a prevalence of short-memory variance structures, similar to the benchmark model. These findings, in contrast to other findings relating to the role of $IM\varepsilon_t$, suggest that the second residual market factor has a notable impact on the structure of the conditional variance. The conditional variance structures of the unrestricted specification now more closely resemble those of the benchmark specification.

As heteroscedasticity in the conditional variance structure can be related to omitted factors, a variance structure that more closely approximates that of the fully specified benchmark model suggests that $IM\varepsilon_t$ reflects influences that are not reflected in $M\varepsilon_t$ (Webster, 2013: 230; Armitage & Brzeszczyński , 2011: 1529). This is evidence in favour of incorporating a second residual market factor and a challenge to the adequacy of the conventional residual market factor. These results favour the incorporation of a second residual market factor, in contrast to other results in this chapter, which are somewhat ambiguous. While the inclusion of two residual market factors yields an improvement in approximating conditional variance structures, both models fail to fully approximate the conditional variance structures underlying the benchmark model.

The mean conditional heteroscedasticity parameters, the mean ARCH coefficients, α_i , for the unrestricted specifications are 0.123 and 0.125 respectively. This is a somewhat surprising finding. As suggested by Bera *et al.* (1988), higher levels of underspecification resulting from factor omission, as in the restricted model, should be associated with higher levels of conditional heteroscedasticity. Therefore, it is surprising that the respective α_i s in Panel C of Table 10.3. are higher than those for the restricted specification (0.109) in Panel C of Table 9.3 and the benchmark specification (0.103) in Panel C of Table 8.3. Differences are not statistically significant.

In measuring the impact of factor omission, it appears that comparisons of the overall conditional variance structures are meaningful whereas comparisons of the level of conditional heteroscedasticity are not (see discussion relating to this point in Section 6.4.6.). A potential explanation for the lack of differences in overall conditional heteroscedasticity relates to the approach employed in this study, that of employing increasingly complex ARCH*(p*) and GARCH(*p,q*) specifications to ensure that the residuals of a specification are free of non-linear dependence and ARCH effects. Bera *et al.* (1988: 204) take a standardised model approach and apply an ARCH(1) specification to their sample of firms and portfolios of CRSP stocks. The use of only two ARCH/GARCH-type specifications in this study may result in a misspecification of the conditional variance model, which will make the ARCH coefficient uninterpretable (Nelson & Cao, 1992). It is plausible that the impact of factor omission on conditional heteroscedasticity is reflected in both ARCH and GARCH coefficients and not just the ARCH parameter as postulated by Bera *et al.* (1988). As a result, quantifying the level of conditional heteroscedasticity when ARCH*(p)* or GARCH(*p,q)* specifications differ (as opposed to a standardised model) across return series may not yield meaningful results. The results are therefore ambiguous and the impact of underspecification on the parameters of more complex ARCH/GARCH specifications (specifications other than the ARCH(p) or GARCH(*p,q*) specifications, such as the EGARCH(*p,q,n*) and IGARCH(*p,q*) specifications) is recommended as an area of further research. The results and the preceding discussion nevertheless suggest that underspecification impacts the overall structure of conditional variance and that the inclusion of residual market factors improves the approximation of conditional variance structures but does not fully result in an approximation of the conditional variance structures associated with the benchmark specification.

In summary, the inclusion of the residual market factors is associated with lower coefficient standard errors and increases in *z*-scores. The largest efficiency gains are from the inclusion of $M\varepsilon_t$ and the contribution of $IM\varepsilon_t$ is marginal in this regard. The inclusion of the conventional residual market factor reduces the upward bias in residual variance and there are also apparent gains from including a second residual market factor when residual variance estimates are compared across specifications on an individual sector basis using the Brown-Forsythe test. Nevertheless, the residual variance estimates are still significantly larger in comparison to the benchmark model. Consequently, the misidentification of the linear factor model, especially at individual sector level, may still pose a challenge to the interpretation, inference making and the testing of the APT relation. The impact of the residual market factors on conditional heteroscedasticity, as measured by the ARCH coefficient, is ambiguous and this may be due to the limitation of relying on different specifications to model the conditional variance structures. In contrast, the conditional variance structures appear to be impacted by factor omission. The inclusion of the residual market factors results in conditional variance structures that more closely approximate those underlying a fully specified model. The second residual market factor, $IM\varepsilon_t$, makes a contribution to approximating the conditional variance structures, in addition to $M\varepsilon_t$.

10.6. PREDICTIVE ABILITY

If the residual market factor is an adequate proxy for omitted factors, then the incorporation of $M\varepsilon_t$ should equate the predictive ability of the unrestricted market model to that of the benchmark model. Mean residuals, ε_{it}, should not differ significantly from zero, as in Panel A of Table 8.4. Panel A of Table 10.4. indicates that the mean residuals of the unrestricted market model are significantly different from zero. This finding is supported by both tests of significance. This result contrasts with the findings of Chang (1991: 387), who reports that the inclusion of a residual market factor in a linear factor model, together with macroeconomic factors, renders mean errors statistically insignificant. However, as in Chang (1991), there is a significant reduction in the magnitude of the mean residuals, indicating that the inclusion of $M\varepsilon_t$ significantly improves the predictive ability of the linear factor model. Residuals decrease (in absolute terms) from 0.0019261 for the restricted specification to 0.0012502 for the unrestricted market model. Nevertheless, the residuals of the unrestricted market model are still more than twice as large as those of the benchmark specification of 0.0005692. This suggests that the inclusion of the residual market factor fails to produce residuals that approximate those of the benchmark specification. Residuals remain overstated and this suggests that the tendency to erroneously not reject the null hypothesis of no factor significance may persist. This is also suggested by the upward bias that persists in the residual variance, as noted in Section 10.5. These findings do not provide support for the adequacy of $M\varepsilon_t$ as a proxy for omitted factors that will resolve concerns relating to factor omission and improve predictive accuracy.

The inclusion of the second residual market factor, $IM\varepsilon_t$, results in a further decrease in the absolute magnitude of the residual errors from 0.0012502 for the unrestricted market model to 0.0010557 for the unrestricted model. The residuals of the unrestricted specification are still almost twice (1.855 fold) as large as those of the benchmark model. Residuals are also significantly different from zero and therefore are not comparable to those of the benchmark specification. The predictive ability of the unrestricted model, like that of the unrestricted market model, underperforms that of the benchmark model. It appears that the inclusion of a second residual market factor has some effect on the magnitude of discrepancies between actual observed returns and predicted returns. However, it is debatable whether the impact is of sufficient magnitude to warrant inclusion in linear factor models as a standard factor in addition to the conventional residual market factor. The minor reduction in the magnitude of the residuals may be responsible for the (slightly) lower number of discrepancies between the total number of statistically significant coefficients in the unrestricted model and the benchmark model relative to the unrestricted market model, as outlined in Section 10.3.1.

Table 10.4: Summary Of Mean Errors And Theil's *U* Statistic For The Unrestricted Models

Panel A: Mean Errors						
	Unrestricted Market Model			**Unrestricted Model**		
	Mean Value			**Mean Value**		
ε_{it}	-0.0012502**▼**▲**			-0.0010557**▼**▲**		
Panel B: Theil's *U* Statistic and Decomposition						
	Mean Value	**Minimum**	**Maximum**	**Mean Value**	**Minimum**	**Maximum**
Theil *U*	0.511▲*▼*	0.316 Mining	0.665 Fixed line telecommunications	0.503▲*▼*	0.301 Mining	0.647 Fixed line telecom.
Bias (U_{BIAS})	0.002002▲W▼	0 Food produces	0.022061 Software & computer services	0.001408▲▼*	0 Food producers Banks	0.021373 Software & computer services
Variance (U_{VAR})	0.269569▲*▼*	0.092939 Mining	0.500836 Software & computer services	0.265393▲*▼*	0.085444 Mining	0.508700 Software & computer services
Covariance (U_{COV})	0.727644▼*▲*	0.477103 Software & computer	0.906429 Mining	0.733199▼*▲*	0.469926 Software & computer services	0.914544 Mining

Notes:
The asterisks, ***, ** and *, indicate statistical significance at the respective 1%, 5% and 10% levels of significance. In Panel A, the Mean Value is the respective arithmetic mean of the residual terms. A paired-sample *t*-test to test the null hypothesis that the mean value of the residuals differs significantly from zero. In Panel B, the Mean Value is the arithmetic mean of the respective measures of predictive accuracy. The Minimum and Maximum are the minimum and maximum values associated with the respective measures of accuracy for the respective sectors. Across panels, the first ▲ or ▼ symbol indicates that a value is larger or smaller relative to that observed for the benchmark model. The second ▲ or ▼ symbol indicates that a value is larger or smaller relative to that observed for the restricted model. Accompanying asterisks, if present, indicate that differences are statistically significant. A superscript "W" indicates a discrepancy between the results of the *t*-test and the Wilcoxon test.

The mean Theil *U* statistics reported for the respective unrestricted specifications in Panel B of Table 10.4. show desirable and statistically significant decreases relative to that of the restricted specification in Panel B of Table 9.4. The mean *U* statistic is 0.511 for the unrestricted market model and 0.503 for the unrestricted specification. This is in contrast to 0.633 for the restricted model. The minimum and maximum values shift downwards relative to those reported for the restricted model. For the unrestricted market model, *U* statistics range between 0.316 for the mining sector and 0.665 for the fixed telecommunications sector. For the unrestricted specification, *U* statistics range between 0.301 for the mining industrial sector and 0.647 for the fixed line telecommunications sector. The lower mean *U* statistics and downward shifts in the minimum and maximum values point towards an improvement in the ability of both models to predict returns. Nevertheless, the mean *U* statistics are significantly higher than that of the benchmark specification of 0.395. The contribution of $IM\varepsilon_t$ to improving predictive accuracy is almost negligible. Both specifications continue to underperform the benchmark specification in predicting returns (Frank, 2009: 58). It appears that a factor analytic augmentation that forms part of the benchmark model is, in addition to the residual market factors, required to incorporate the influence of omitted factors that are not reflected in the two residual market factors.

The comparisons of the mean bias proportion, U_{BIAS}, for the unrestricted specifications to those of the restricted and benchmark models are somewhat ambiguous. The inclusion of the residual market factor decreases the mean bias proportion from 0.002355 for the restricted specification in Panel B of Table 9.4. to 0.002002 for the unrestricted market model. The difference is not statistically significant. This suggests that the mean bias proportions of the restricted model and the unrestricted market model are comparable. The paired-sample *t*-test also indicates that the bias proportion does not differ significantly from that of the benchmark model of 0.001036. However, this is contradicted by the results of the Wilcoxon matched-pairs signed-rank test that suggests that the bias proportion of the unrestricted market model is significantly higher than that of the benchmark specification. It appears that the inclusion of $M\varepsilon_t$ yields somewhat ambiguous and potentially non-existent improvements. The bias proportion for the unrestricted market model remains below 0.1 but is still higher relative to that of the benchmark specification. The inclusion of $IM\varepsilon_t$ results in a reduction in the mean bias proportion to an extent that is more closely comparable to that of the benchmark specification. The mean bias proportion is now 0.001408 and the differences between the bias proportions of the unrestricted model and benchmark model are statistically insignificant. The mean bias proportion is also significantly lower than that of the restricted specification of 0.002355. This suggests that $M\varepsilon_t$ by itself does not produce substantial reductions in bias and consequent reductions in the systematic error of predicted values.

The bias proportion for the unrestricted market model ranges between zero for the food producers sector and 0.022061 for the software and computer services sector. Interestingly, two sectors exhibit a bias proportion of zero for the unrestricted model; food producers and banks. The sector with a maximum bias proportion of 0.021373 is the software and computer services sector. Also interestingly, the maximum bias proportion value for the unrestricted model is higher than that of the restricted model, with the software and computer services sector exhibiting a bias proportion of 0.020372. Nevertheless, the mean bias proportions of the unrestricted models are still lower than that of the restricted model. The minimum values for both unrestricted specifications are comparable to those of the benchmark model of zero for the industrial metals and mining sector.

The impact on the variance proportion, U_{VAR} , the measure of the ability of a specification to replicate the variance of the actual observations, is more pronounced (Kacapyr, 2014: 162). The mean variance proportion decreases from 0.412575 for the restricted specification in Panel B of Table 9.4. to 0.269569 and 0.265393 in Panel B of Table 10.4. (an approximate decrease of 34.662% in both instances) for the respective unrestricted specifications. Differences between the variance proportions of the restricted model and the unrestricted specifications are statistically significant. Such reductions are desirable and indicate that $M\varepsilon_t$ by itself improves the ability of the linear factor model to replicate the variance of the actual return series. The contribution of $IM\varepsilon_t$ is negligible. The mean variance proportions of both unrestricted specifications continue to significantly exceed that of the benchmark specification of 0.178215 (approximately 33.889% lower) in Panel B of Table 8.4. This implies that both residual market factors do not encompass information that is required to accurately replicate the variance of the actual series.

A comparison of the respective minimum and maximum values confirms this. For the unrestricted market model, the mining sector has a variance proportion of 0.092939 and the software and computer services sector has a variance proportion of 0.500836. For the unrestricted model, the mining sector has a variance proportion of 0.085444 and the software and computer services sector has a variance proportion of 0.508700. The minimum variance proportion values are higher relative to those of the benchmark model for which the lowest variance proportion value is 0.022184, for the mining sector. The maximum values are comparable to those of the benchmark model; the software and computer services sector has a variance proportion of 0.509499 (Panel B of Table 8.4.). The extreme values, notably the minimums, are lower relative to those of the restricted model. In the restricted model in Panel B of Table 9.4., the industrial transport sector has a minimum variance proportion of 0.287454 whereas the software and computer services sector has a maximum variance proportion of 0.613429. In terms of predicting the variance of a return series, the bias persists and is not resolved by the incorporation of $M\varepsilon_t$ and also $IM\varepsilon_t$. The inclusion of the conventional residual market factor nevertheless leads to a significant reduction in the mean variance proportion relative to the restricted model. The contribution of $IM\varepsilon_t$ is minimal.

The mean covariance proportion, U_{COV}, is 0.807705 for the benchmark specification in Panel B of Table 8.4. whereas the covariance proportions for the unrestricted market model and the unrestricted model in Panel B of Table 10.4. are 0.727644 and 0.733199 respectively. The almost non-existent decrease in the mean covariance proportions for the unrestricted models from 0.727644 to 0.733199 suggests that the inclusion of $IM\varepsilon_t$ has no impact on reducing systematic error (Brooks & Tsolacos, 2010: 272). Even so, the mean covariance proportion measures are still significantly lower relative to that of the benchmark specification. This suggests that a greater proportion of the prediction error in these models is attributable to the structure of the respective specifications rather than inherent randomness of the data (Fildes & Kingsman, 2011: 487). In contrast, the mean covariance proportion in Panel B of Table 9.4. for the restricted specification is 0.585071, approximately 19.594% and 20.203% lower than that of the respective unrestricted specifications. Differences between the mean covariance proportions of the unrestricted models and the restricted model are statistically significant. This suggests that the inclusion of $M\varepsilon_t$ significantly lowers systematic prediction errors and a greater proportion of prediction error is now attributable to the inherent randomness of the data (Fildes & Kingsman, 2011: 487). The use of two residual market factors fails overall to reduce the proportion of prediction errors attributable to systematic errors to an extent that is comparable to that of the benchmark model.

Minimum and maximum covariance proportion values confirm the overall findings. For the unrestricted market model, the minimum value is 0.477103 for the software and computer services sector and the maximum value is 0.906429 for the mining sector. For the unrestricted model, the software and computer services sector is associated with a covariance proportion of 0.469926 whereas the mining sector is associated with a covariance proportion of 0.914544. The minimum and maximum values for both specifications are greater than those for the restricted model of 0.366200 for the software and computer services sector and 0.711488 for the industrial transport sector, respectively. The minimum values approximate that of the benchmark model of 0.472995 for the software and computer services sector but are slightly lower than the maximum value of 0.976984 for the mining sector.

The results in Panel A of Table 10.4. indicate that the inclusion of the residual market factors, most notably $M\varepsilon_t$, reduces prediction errors. Mean errors are nevertheless still significantly different from zero and this is in contrast to the results of the benchmark model. The inclusion of the residual market factors improves predictive performance relative to the restricted model, as indicated by lower mean *U* statistics. The inclusion of both residual market factors also improves the ability of models to replicate the variability of the return series although the greatest gains are associated with the inclusion of $M\varepsilon_t$. A greater proportion of the prediction error can now be attributed to inherent randomness in the data, as opposed to the structure of the return generating process, relative to the restricted specification. Although residual market factors improve predictive ability, specifications that combine macroeconomic factors with a residual factor or factors will still produce suboptimal predictions. The contribution of a second residual market factor to optimising predictive ability is almost negligible, with the exception of its apparent contribution to reducing systematic bias.

10.7. FACTOR OMISSION

The results of the LR test carried out for the unrestricted specifications are summarised in Panel A of Table 10.5. As with the restricted model, the results suggest that the unrestricted models are underspecified.

Table 10.5: Likelihood Ratio Test And Factor Analysis Summary For The Unrestricted Models

Unrestricted Market Model			Unrestricted Model		
Panel A: Likelihood Ratio Test For Omitted Factors					
Omitted Factor	**Mean LR Statistic**	**Total Sig.**	**Mean LR Statistic**	**Total Sig.**	
$IM\varepsilon_t$	4.920	13/26	-	-	
f_{1t}, f_{2t} ,	64.631	25/26	73.730	25/26	
Panel B: Full Period Factor Analysis					
Extracted factor(s)	**Mean Communality**	**Mean Uniqueness**	**Extracted factor(s)**	**Mean Communality**	**Mean Uniqueness**
3	0.300▲▼	0.700▼▲	2	0.248▲▼	0.752▼▲
Panel C: Subperiod Factor Analysis					
Period: 2001M016 to 2008M12					
2	0.326▲▼	0.674▼▲	2	0.309▲▼	0.691▼▲
Period: 2009M01 to 2016M12					
3	0.284▲▼	0.716▼▲	3	0.286▲▼	0.714▼▲

Notes:
In Panel A, the Mean LR Statistic is the mean of the LR test statistics from the Likelihood Ratio test for omitted factors. Total Sig. is the number of outcomes rejecting the null hypothesis that a given factor or set of factors has not been omitted. Significance is recorded at the 10% level of significance. In Panel B and Panel C, Mean Communality is the mean proportion of common variance explained across return series by common factors extracted on the basis of the MAP test. Mean Uniqueness is the mean proportion of variance across return series attributable to the return series themselves and not the systematic factors. Across panels, the first ▲ or ▼ symbol indicates that a value is larger or smaller relative to that observed for the benchmark model. The second ▲ or ▼ symbol indicates that a value is larger or smaller relative to that observed for the restricted model. Accompanying asterisks, if present, indicate that differences are statistically significant. The superscript "W" indicates that the Wilcoxon matched-pairs signed-rank test contradicts the results of the paired-sample *t*-test.

For the unrestricted market model, the hypothesis that $IM\varepsilon_t$ is insignificant is rejected in 13 out of 26 instances. This confirms that as with the restricted specification, the macroeconomic factor set and the residual market factor fail to account for international influences in returns (Section 4.2.2.; Panel A of Table 9.5.). Moreover, the results also confirm the existence of unidentified and unobserved influences in returns. The null hypothesis that the factor analytic augmentation is statistically insignificant is rejected for all sectors, with the exception of the forestry and paper sector. For the unrestricted model, the null hypothesis of the insignificance of the factor analytic factors is also rejected for all sectors with the exception of the forestry and paper sector. This again confirms the presence of unobserved factors in returns that are not accounted for by the macroeconomic factor set and both $M\varepsilon_t$ and $IM\varepsilon_t$ (Section 7.4.; Section 8.3.2.). These results, for both unrestricted specifications, as well as those for the restricted model in Panel A of Table 9.5., confirm the results in Panel A of Table 8.1. relating to the widespread signficance of the statistically derived factors in the benchmark model. Both unrestricted models continue to suffer from underspecification.

The scree plots in Figure 10.1. suggest that there are two common factors in the residuals of both unrestricted specifications over the full sample period. This finding is comparable with that of the scree plot for the residuals of the restricted specification in Figure 9.1. A visual comparison of both scree plots in Figure 10.1. to Figure 9.1. shows less distinct flexion points for both unrestricted specifications. The flexion points are nevertheless more distinct relative to that of the scree plot for the residuals of the benchmark specification, which is inconclusive.[152]

The MAP test confirms the presence of multiple factors in the residuals of the unrestricted specifications. Three factors are extracted from the residuals of the unrestricted market model and two factors are extracted from the residuals of the unrestricted model. The respective mean communalities associated with these factors are 0.300 and 0.248 implying that omitted factors continue to explain a substantial amount of shared variance (30.0% and 24.8% respectively) (Panel B of Table 10.5.). The decrease in the mean communality from 0.300 for the unrestricted market model to 0.248 for the unrestricted model suggests that

[152] The scree plot for the benchmark model does not exhibit a distinct flexion point and is not reported in Section 8.7.

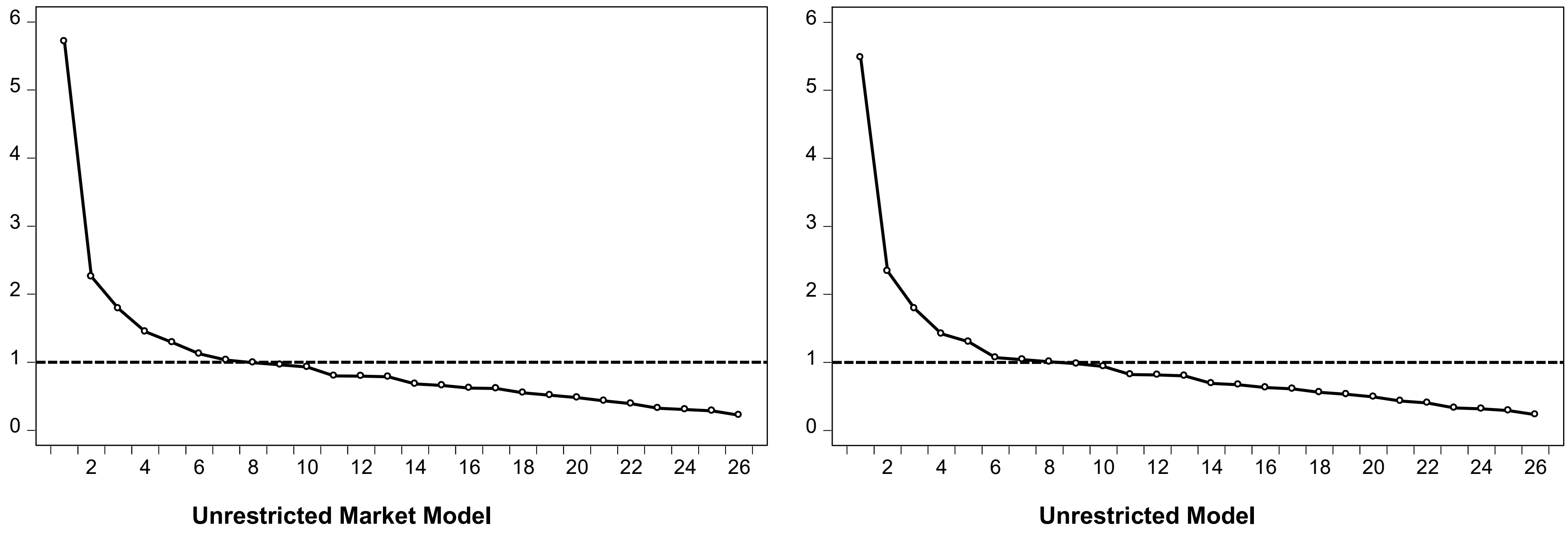
there are some pervasive and global influences in $IM\varepsilon_t$ that are not reflected by $M\varepsilon_t$. Mean uniqueness values are 0.700 and 0.752 respectively.

These results contrast starkly with the shared variance of 6.6%, which is explained by the single common factor extracted from the residuals of the benchmark specification and with a mean uniqueness of 0.934, as reported in Panel A of Table 8.5. These results, for the unrestricted specifications, are nevertheless an improvement over the restricted specification. The three factors extracted from the residuals of the restricted specification, on average, explain 39.9% of shared variance. Mean uniqueness is 0.601 (Panel B of Table 9.5.). These results are expected; the macroeconomic and residual market factors in the unrestricted specifications explain a greater (lower) proportion of common variation in return series relative to the restricted (benchmark) specification.

These results indicate that although the conventional residual market factor accounts for some variation in the residuals associated with omitted factors, a substantial amount of variation can be attributed to omitted pervasive factors that are not reflected by the residual market factor. Consequently, the inclusion of the second residual market factor in the unrestricted specification appears to contribute to capturing common variation in the residuals. That $IM\varepsilon_t$ reflects influences not reflected in $M\varepsilon_t$ is also suggested by the lower number of factors extracted from the residuals of the unrestricted model. This is further suggested by the decline in the mean communality from 0.300 for the unrestricted market model to 0.248 for the unrestricted model. For the unrestricted market model, 23 sectors are associated with communalities above 0.15 and for the unrestricted specification, 19 sectors are associated with communalities above 0.15. This is in contrast with five sectors for the benchmark specification, again suggesting that the residual market factors fail to reduce common variation to the same extent across individual sectors as the factor set in the benchmark specification. As with other results, the use of residual market factors results in an improvement over the restricted specification for which all residual series are associated with communalities above 0.15.[153]

[153] The unabridged results of this factor analysis are available upon request.

Figure 10.1: Scree Plot Of Eigenvalues For Residuals Of The Unrestricted Models



Next, the MAP test is applied to factor analyse the residual correlation matrices of the unrestricted market model and the unrestricted model over the 2001M01 to 2008M12 and 2009M01 to 2016M12 periods. Results are reported in Panel C of Table 10.5. For the 2001M01 to 2008M12 period, two factors are extracted for both unrestricted specifications with respective mean communalities of 0.326 and 0.309. The respective mean uniqueness measures are 0.674 and 0.691 for this subperiod. For the 2009M01 to 2016M12 subperiod, three factors are extracted from the residuals of each unrestricted specification with respective mean communalities of 0.284 and 0.286. Mean uniqueness is 0.716 and 0.714 respectively. These results suggest that there are most likely two common factors in the residuals that are of a systematic nature. The third factor is potentially a transient pseudofactor, which emerges during the second half of the sample period.

Of concern is that the factors extracted for both specifications over the 2009M01 and 2016M12 period have mean communalities close to that of the factors extracted from the residuals of the restricted specification over this period, with a mean communality of 0.339 (Panel C of Table 9.5.). This suggests that the residual market factors are more effective during the first half of the sample period, which is associated with a mean communality of 0.461 for the restricted model but with substantially lower mean communalities of 0.326 and 0.309 for the respective unrestricted models. For the benchmark specification, no factors are extracted over the 2001M01 to 2008M12 period and the factor extracted over the 2009M01 to 2016M12 period resembles a pseudofactor, given that its existence is confined to the second half of the sample period.

The results of the factor analysis suggest that although the residual market factors are proxies for omitted factors, they are inadequate proxies. Statistical factors, representative of pervasive influences in returns, can be extracted from the respective residual correlation matrices. The extracted factor scores explain a significant proportion of shared variance, which is not accounted for by the macroeconomic factor set and the residual market factors. There are at least two such factors in the residuals of both unrestricted specifications and these factors do not appear to be of a transitory nature.

Residual market factors are designated as proxies for omitted factors in the literature. Yet, the emergence of factors from the respective residual correlation matrices of the unrestricted models suggests that the adequacy of residual market factors as proxies for omitted factors needs to be interrogated. The overall validity of such models , of models that combine macroeconomic factors with residual market factors to resolve underspecification, is seemingly questionable (Meyers, 1973: 698; 705). Within this context, the efficacy of the approach of incorporating residual market factors to proxy for omitted factors is also questionable. Finally, the mean communality of the factors extracted over the entire sample period (reported in Panel C of Table 10.5.) is lower for the unrestricted model relative to the unrestricted market model. This suggests that $IM\varepsilon_t$ captures some of the pervasive influences in returns that are not reflected in $M\varepsilon_t$. If $M\varepsilon_t$ is a true catch-all proxy, all influences should be reflected in the conventional residual market factor and mean communalities should be similar for both unrestricted specifications. This does not seem to be the case.

10.8. THE RESIDUAL CORRELATION MATRIX

APT literature that focuses on the linear factor model proposes that a residual market factor will adequately account for omitted factors (Section 3.2.; Section 3.4.). The results presented in this chapter, and specifically those of the LR tests for omitted factors and those of factor analysis, point towards the omission of factors and the existence of pervasive factors that are reflected in the respective residual correlation matrices. Furthermore, the inclusion of a second residual market factor, $IM\varepsilon_t$, does not appear to significantly alleviate the situation.

The histograms in Figure 10.2. reproduce the respective residual correlation matrices for the two unrestricted specifications. Correlation coefficients are summarised in Table 10.6.:

Figure 10.2: Histogram of Unrestricted Model Residual Correlation Coefficients

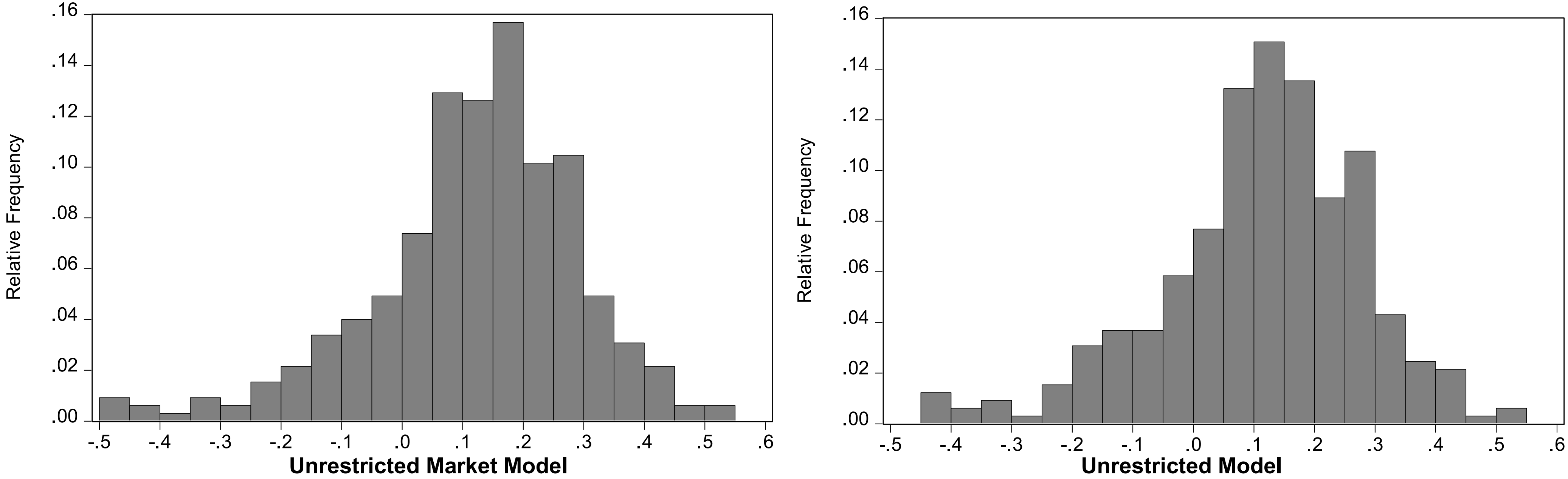


Table 10.6. Distribution Summary Of The Unrestricted Model Residual Correlation Matrix

	Unrestricted Market Model			Unrestricted Market Model		
Bin	**Frequency**	**Relative Frequency**	**Cumulative Frequency**	**Frequency**	**Relative Frequency**	**Cumulative Frequency**
$-0.5 < \rho_{ij} \leq -0.4$	5	1.538%	1.538%	4	1.231%	1.231%
$-0.4 < \rho_{ij} \leq -0.3$	4	1.231%	2.769%	5	1.538%	2.769%
$-0.3 < \rho_{ij} \leq -0.2$	7	2.154%	4.923%	6	1.846%	4.615%
$-0.2 < \rho_{ij} \leq -0.1$	18	5.538%	10.462%	22	6.769%	11.385%
$-0.1 < \rho_{ij} \leq 0$	29	8.923%	19.385%	31	9.538%	20.923%
$0 < \rho_{ij} \leq 0.1$	66	20.308%	36.692%	68	20.923%	41.846%
$0.1 < \rho_{ij} \leq 0.2$	92	28.308%	60.000%	93	28.615%	70.462%
$0.2 < \rho_{ij} \leq 0.3$	67	20.615%	88.615%	64	19.692%	90.154%
$0.3 < \rho_{ij} \leq 0.4$	26	8.000%	96.615%	22	6.769%	96.923%
$0.4 < \rho_{ij} \leq 0.5$	9	2.769%	99.385%	8	2.462%	99.385%
$0.5 < \rho_{ij} \leq 0.6$	2	0.615%	100.00%	2	0.615%	100.00%
$0.6 < \rho_{ij} \leq 0.7$	0	0	100.00%	0	0.00%	100.00%
Total	**325**	**100.00%**	**100.00%**	325	**100.00%**	**100.00%**
Mean	0.120***▲***▼***			0.111***▲***▼***		
Minimum	-0.498			-0.448		
Maximum	0.543			0.543		

Notes:
The asterisks, ***, ** and *, indicate statistical significance at the respective 1%, 5% and 10% levels of significance. The *t*-test is applied to test the hypothesis that the mean of correlation coefficients does not differ significantly from zero. The Wilcoxon matched-pairs signed-rank test is applied as a confirmatory test and the superscript "W" indicates that the Wilcoxon matched-pairs signed-rank test contradicts the results of the paired-sample *t*-test. . Bin represents ranges of correlation coefficients and frequency reports the number of correlation coefficients that fall within each range. Relative Frequency is the percentage of correlation coefficients that fall within the respective ranges. Cumulative Frequency is the running total of all previous relative frequencies.
Mean is the mean value of correlation coefficients in the correlation matrix and the Minimum and Maximum are the lowest and largest correlation coefficients observed. Across panels, the first ▲ or ▼ symbol indicates that a value is smaller or larger relative to that observed for the benchmark model. Accompanying asterisks, if present, indicate that differences are statistically significant. The second ▲ or ▼ symbol indicates that a value is larger or smaller relative to that observed for the restricted model.

An analysis of the histograms of the residual correlation coefficients indicates that in general, the magnitude of correlation coefficients has decreased relative to that of the restricted specification but is still greater than that of the benchmark model (see Figure 9.2; Figure 8.1). The histogram of the residual correlation coefficients derived from the residuals of the restricted model indicates that the vast majority of correlation coefficients are positive and that observations are centred around 0.3. The histograms for the unrestricted market model and for the unrestricted model more closely resemble those of the benchmark specification. Most correlation coefficients are centred around 0.1. This is especially evident from the histogram of correlation coefficients for the unrestricted model. In contrast to the histogram of the correlation coefficients in Figure 9.2., most correlation coefficients are substantially lower than 0.3 for the unrestricted specifications. For the unrestricted market model, 60.231% (225 out of 325) of correlation coefficients are greater than zero but less or equal to 0.3. Importantly, 31.385% (102 of 325) of correlation coefficients lie within the -0.12 and 0.12 range and are likely to be statistically insignificant. Although this suggests a substantially greater number of correlation coefficients of a low magnitude relative to the restricted specification for which 7.077% (23 of 325) of correlation coefficients fall within this range, this percentage (and number) is still far below that of the benchmark specification of 71.629% (233 of 325) (Section 8.8.; Section 9.8.). The inclusion of $IM\varepsilon_t$ yields some improvement. When $IM\varepsilon_t$ is incorporated into the unrestricted specification, 37.538% (122 of 325) of correlation coefficients are of a low magnitude, falling within the -0.12 and 0.12 range.

The significant decrease in the mean level of correlation, from 0.291 in Table 9.6. to 0.120 in Table 10.6 for the unrestricted market model again points towards the contribution of the residual market factor, $M\varepsilon_t$, to accounting for remaining co-movement in returns. The marginally lower mean residual correlation for the unrestricted model of 0.111 again suggests that the contribution of the second residual market factor is minor. In both instances, the mean level of correlation is significantly higher than that of the residual correlation matrix of the benchmark model (in absolute terms, 0.024; Table 8.6.), pointing towards the presence of omitted pervasive factors in the residuals.

The minimum and maximum values for both specifications reflect the general decrease in the level of correlation. The respective minimum and maximum values for the unrestricted market model range between -0.498 and 0.543 and between -0.448 and 0.543 for the unrestricted specification. The range of residual correlation coefficients for the benchmark specification is noticeably narrower, with correlation ranging between -0.320 and 0.396 (Table 8.5.). For the residuals of the restricted model, the lower limit is -0.003 and the upper limit is 0.615. The range of correlation coefficients for this model exhibits higher minimum and maximum values relative to those obtained from the unrestricted models. This indicates that the inclusion of the residual market factors translates into a residual correlation structure which more closely resembles that of the benchmark specification but does not approximate it fully. Given that the residual correlation structure of the unrestricted specifications differs from that of the benchmark specification and given that the mean level of correlation is significantly higher than that derived from the residuals of the benchmark model, the residual market factors do not appear to adequately account for omitted factors that are responsible for remaining residual interdependence (Elton *et al.,* 2014: 157).

The residual correlation matrices of the respective unrestricted models are reproduced in Table 10.7. and Table 10.8. It is immediately apparent that the number of individual statistically significant coefficients (shaded pairwise correlation coefficients are statistically significant) for the unrestricted specifications is lower relative to that of the restricted specification, as reported in Table 9.7. A comparison of the respective correlation matrices in Table 10.7. and Table 10.8. confirms that the inclusion of a second residual market factor has a marginal impact on the structure of the residual correlation matrix. A visual examination shows that the difference between the number of statistically significant correlation coefficients is almost indiscernible. In contrast to the matrix derived from the benchmark specification which exhibits mostly unsystematic correlation within economic sectors and outside of economic sectors, the matrices in Table 10.7. and Table 10.8. exhibit widespread positive residual correlation within economic sectors and outside of the respective economic sector submatrices. This is indicative of co-movement in a common direction, potentially attributable to the omission of unspecified common factors that are relegated to the residuals but comprise the factor analytic augmentation in the benchmark model.

Table 10.7: Correlation Matrix Of The Unrestricted Market Model Residuals

	J135	J173	J175	J177	J235	J272	J273	J275	J277	J279	J335	J353	J357	J453	J457	J533	J537	J555	J575	J653	J835	J853	J857	J877	J898	J953
J135	1.000																									
J173	0.083	1.000																								
J175	0.052	-0.074	1.000																							
J177	-0.152	-0.088	0.449	1.000																						
J235	0.314	0.030	0.225	-0.076	1.000																					
J272	0.093	-0.042	-0.051	-0.387	0.250	1.000																				
J273	0.337	0.039	0.055	-0.188	0.413	0.216	1.000																			
J275	0.239	0.118	0.124	-0.006	0.213	-0.005	0.373	1.000																		
J277	0.324	0.004	0.015	-0.216	0.271	0.189	0.278	0.296	1.000																	
J279	0.236	0.111	-0.045	-0.335	0.332	0.360	0.269	0.054	0.361	1.000																
J335	0.291	0.075	0.098	-0.083	0.067	0.078	0.180	0.189	0.217	0.138	1.000															
J353	0.054	0.140	-0.168	-0.334	-0.117	0.121	0.124	0.059	-0.033	-0.021	0.047	1.000														
J357	0.154	-0.131	0.029	-0.223	0.151	0.274	0.267	0.052	0.298	0.255	0.138	0.133	1.000													
J453	0.239	0.019	-0.049	-0.193	0.176	0.230	0.201	0.093	0.288	0.183	0.079	0.152	0.197	1.000												
J457	0.158	-0.077	0.044	-0.196	0.121	0.072	0.205	0.047	0.149	0.175	0.162	0.033	0.205	0.271	1.000											
J533	0.115	-0.124	0.011	-0.285	0.196	0.397	0.126	0.012	0.310	0.243	0.085	0.003	0.353	0.235	0.232	1.000										
J537	0.162	-0.135	-0.085	-0.467	0.237	0.424	0.321	0.103	0.457	0.429	0.209	-0.019	0.405	0.294	0.374	0.543	1.000									
J555	0.187	-0.093	-0.217	-0.407	0.048	0.260	0.114	-0.001	0.263	0.296	0.071	0.078	0.094	0.164	0.151	0.129	0.328	1.000								
J575	0.293	0.050	-0.015	-0.263	0.302	0.193	0.214	0.291	0.399	0.368	0.173	0.132	0.267	0.253	0.159	0.138	0.286	0.205	1.000							
J653	0.107	-0.147	-0.129	-0.145	0.011	0.117	0.082	-0.024	0.159	0.156	0.079	-0.181	0.117	0.004	0.180	0.132	0.242	0.269	0.115	1.000						
J835	0.062	-0.144	-0.030	-0.434	0.155	0.392	0.169	0.067	0.297	0.273	0.012	-0.087	0.236	0.166	0.219	0.324	0.517	0.182	0.220	0.138	1.000					
J853	0.157	-0.106	0.106	-0.138	0.168	0.179	0.185	0.078	0.306	0.099	0.102	0.074	0.296	0.265	0.134	0.272	0.222	0.002	0.186	0.072	0.172	1.000				
J857	0.186	0.043	-0.111	-0.498	0.034	0.187	0.146	0.132	0.174	0.218	0.091	0.079	0.180	0.167	0.195	0.244	0.319	0.198	0.204	0.098	0.460	0.195	1.000			
J877	0.226	-0.068	-0.220	-0.462	0.136	0.315	0.281	-0.038	0.155	0.313	0.077	0.134	0.208	0.022	0.143	0.295	0.427	0.319	0.252	0.200	0.448	0.051	0.390	1.000		
J898	0.184	0.076	-0.173	-0.313	0.099	0.178	0.132	-0.040	0.140	0.214	0.155	0.286	0.177	0.115	0.148	0.148	0.212	0.205	0.216	0.091	0.065	0.176	0.300	0.268	1.000	
J953	0.017	-0.048	-0.095	-0.210	-0.020	0.055	0.097	-0.062	0.122	0.090	0.094	-0.077	0.059	0.004	0.152	0.069	0.135	0.301	0.172	0.217	0.122	0.181	0.267	0.274	0.135	1.000

Table 10.8: Correlation Matrix Of The Unrestricted Model Residuals

	J135	J173	J175	J177	J235	J272	J273	J275	J277	J279	J335	J353	J357	J453	J457	J533	J537	J555	J575	J653	J835	J853	J857	J877	J898	J953
J135	1.000																									
J173	0.051	1.000																								
J175	0.037	-0.085	1.000																							
J177	-0.107	-0.039	0.486	1.000																						
J235	0.298	0.008	0.219	-0.043	1.000																					
J272	0.090	-0.048	-0.053	-0.392	0.247	1.000																				
J273	0.308	0.001	0.047	-0.135	0.399	0.215	1.000																			
J275	0.210	0.087	0.118	0.051	0.194	-0.011	0.347	1.000																		
J277	0.313	-0.022	0.013	-0.186	0.259	0.186	0.256	0.278	1.000																	
J279	0.215	0.091	-0.053	-0.312	0.320	0.359	0.249	0.030	0.348	1.000																
J335	0.283	0.060	0.094	-0.060	0.056	0.077	0.165	0.175	0.208	0.130	1.000															
J353	0.054	0.139	-0.171	-0.344	-0.119	0.121	0.122	0.056	-0.034	-0.023	0.047	1.000														
J357	0.160	-0.131	0.030	-0.236	0.153	0.275	0.276	0.056	0.303	0.260	0.140	0.133	1.000													
J453	0.255	0.028	-0.046	-0.218	0.182	0.232	0.216	0.104	0.298	0.193	0.084	0.153	0.197	1.000												
J457	0.148	-0.093	0.039	-0.182	0.112	0.070	0.192	0.032	0.142	0.168	0.156	0.033	0.207	0.277	1.000											
J533	0.112	-0.135	0.009	-0.283	0.191	0.396	0.119	0.003	0.308	0.241	0.082	0.003	0.354	0.239	0.230	1.000										
J537	0.138	-0.165	-0.094	-0.448	0.223	0.424	0.299	0.078	0.446	0.419	0.199	-0.022	0.411	0.305	0.367	0.543	1.000									
J555	0.167	-0.119	-0.227	-0.389	0.034	0.259	0.088	-0.027	0.250	0.283	0.059	0.077	0.097	0.172	0.142	0.124	0.315	1.000								
J575	0.266	0.011	-0.023	-0.210	0.284	0.190	0.177	0.261	0.380	0.351	0.157	0.133	0.276	0.271	0.146	0.131	0.263	0.184	1.000							
J653	0.083	-0.175	-0.137	-0.112	-0.006	0.114	0.056	-0.050	0.143	0.142	0.067	-0.186	0.120	0.010	0.171	0.127	0.227	0.255	0.089	1.000						
J835	0.028	-0.181	-0.038	-0.410	0.136	0.395	0.137	0.034	0.279	0.255	0.000	-0.091	0.242	0.179	0.212	0.325	0.505	0.164	0.192	0.119	1.000					
J853	0.155	-0.114	0.104	-0.134	0.165	0.178	0.182	0.073	0.306	0.096	0.100	0.074	0.297	0.267	0.133	0.271	0.220	-0.002	0.184	0.068	0.172	1.000				
J857	0.115	-0.032	-0.161	-0.440	-0.017	0.190	0.067	0.062	0.139	0.181	0.061	0.080	0.202	0.203	0.178	0.246	0.287	0.158	0.133	0.049	0.433	0.197	1.000			
J877	0.186	-0.111	-0.246	-0.425	0.112	0.313	0.245	-0.083	0.128	0.290	0.058	0.133	0.215	0.033	0.128	0.291	0.408	0.300	0.212	0.177	0.424	0.044	0.332	1.000		
J898	0.175	0.061	-0.184	-0.302	0.090	0.171	0.116	-0.058	0.131	0.204	0.146	0.287	0.179	0.122	0.142	0.142	0.202	0.201	0.200	0.080	0.054	0.173	0.288	0.254	1.000	
J953	-0.006	-0.075	-0.103	-0.180	-0.037	0.052	0.066	-0.092	0.107	0.074	0.084	-0.076	0.062	0.013	0.144	0.065	0.115	0.288	0.143	0.198	0.106	0.179	0.235	0.256	0.130	1.000

Support for this hypothesis, namely that of the factor analytic augmentation representing omitted factors, is provided by the results of the LR tests for omitted factors, which indicate that f_1 and f_2 are omitted (and unspecified) factors in both the unrestricted models (Section 10.7.).

Table 10.9: Tests Of Matrix Equality For The Unrestricted Models

Hypothesis	χ^2 Statistic	Reject
	Unrestricted Market Model	
$M_{26} = A_{26}$	783.637***	Reject
$M_{26} = I_{26}$	2769.57***	Reject
$M_{26} = R_{26}$	614.529***	Reject
$M_{26} = B_{26}$	680.222***	Reject
	Unrestricted Model	
$U_{26} = A_{26}$	713.475***	Reject
$U_{26} = I_{26}$	2569.31***	Reject
$U_{26} = R_{26}$	555.979***	Reject
$U_{26} = B_{26}$	663.927***	Reject

Notes:
The asterisks, ***, ** and *, indicate statistical significance at the respective 1%, 5% and 10% levels of significance. Hypothesis is the hypothesis that is being tested relating to the equality of two matrices. χ^2 Statistic is the resultant test statistic for the Jennrich test and Reject indicates whether the null hypothesis of equality between two matrices is rejected. B_{26} denotes the residual correlation matrix derived from the benchmark model. A_{26} denotes the residual correlation matrix of the actual return series. I_{26} denotes the identity matrix. R_{26} denotes the residual correlation matrix derived from the restricted model. M_{26} denotes the residual correlation matrix derived from the unrestricted market model. U_{26} denotes the residual correlation matrix derived from the unrestricted model.

The results of the Jennrich (1970) test, reported in Table 10.9., indicate that the null hypothesis of equality between M_{26} , the unrestricted market model residual correlation matrix, and U_{26}, the unrestricted model residual correlation matrix, and the correlation structure of the actual return series, A_{26}, may be rejected. This confirms the ability of the residual market factors to proxy for pervasive factors in returns, in contrast to the set of macroeconomic factors in the restricted model (Table 9.7.). The null hypothesis of equality between M_{26} and U_{26} and the identity matrix, I_{26}, is also rejected. The existence of omitted factors in the residuals of the respective unrestricted models, represented by the factor analytic augmentation, will translate into stronger residual interdependence (as evident from the results in Section 10.7. and the preceding analysis). Therefore, this result is expected given that this null hypothesis is also rejected for the benchmark model residual correlation matrix, B_{26}. The null hypothesis of equality between M_{26} and U_{26} and the residual correlation matrix derived from the restricted model, R_{26}, is rejected in both instances. This is in line with the preceding findings in this section, which show that the mean level of correlation and the number of significant correlations in the respective correlation matrices decreases. The inclusion of the residual market factors therefore appears to weaken residual interdependence. This is expected if these factors proxy for omitted factors although the contribution of $IM\varepsilon_t$ is marginal and most of the reduction in interdependence is attributable to $M\varepsilon_t$.

Finally, to determine whether the residual market factors adequately account for omitted factors and whether the residual correlation structure is comparable to that of the benchmark model, the equality of M_{26} and U_{26} is tested against that of the benchmark model, B_{26}. The null hypothesis is rejected implying that the use of a single or two residual market factors fails to replicate the residual correlation matrix derived from the benchmark specification. Such a finding implies that there is still significant co-movement that is higher than that in the residual correlation matrix derived from the benchmark model. Moreover, it appears that an approach that not only relies upon one residual market factor but on two residual market factors does not capture all omitted factors.

The findings in this section again directly challenge the validity of relying upon the conventional residual market factor to account for all omitted factors. While the residual market factor reduces pairwise residual correlation, it fails to reduce the levels of residual correlation to such an extent that is it comparable to that of the benchmark model. The contribution of the second residual market factor is marginal and it appears that the underspecification problem will continue to persist unless a factor analytic augmentation approach is followed.

10.9. CHAPTER SUMMARY AND CONCLUSION

The residual market factor is widely employed in the APT literature as a hypothesised summary measure of various omitted and unobserved factors (see Chapter 3). In this chapter, the residual market factor is incorporated into the linear factor model alongside a set of macroeconomic factors. The relevance of a second residual market factor, derived from an international market index, is also considered (see Chapter 4). The resultant specifications - the unrestricted market model and the unrestricted model - are juxtaposed against the benchmark model outlined in Chapter 8 and the restricted model outlined in Chapter 9. The premise of the comparisons and analysis that follow is that if the conventional residual market factor is an adequate proxy for omitted factors, then the unrestricted market model should perform similarly to the benchmark model in numerous aspects. The findings are that this is not the case. Also, a second residual market factor does not contribute significantly to resolving biases associated with factor omission. This is especially evident from the results of the unrestricted model reported in Section 10.3.1. that show that $IM\varepsilon_t$ features prominently in this specification and is not redundant. This suggests that there are other influences in returns that are not fully reflected in $M\varepsilon_t$. Nevertheless, improvements in various aspects, relative to the unrestricted market model, are often minor if not negligible with the inclusion of $IM\varepsilon_t$.

The explanatory power of the unrestricted market model remains below that of the benchmark specification and the significance of factors remains understated. The unrestricted market model continues to underperform in terms of predictive accuracy and deviates from the true return generating process to a greater extent relative to the benchmark model. There are improvements relative to the restricted specification. All factors that are now identified as systematic, including $LEAD_{t-1}$ and $USD\varepsilon_t$, as in the benchmark model. The overall level of understatement is reduced. The inclusion of the second residual market factor reduces understatement further, although the reduction is marginal (Section 10.3.1.; Section 10.3.3.).

Encouragingly, model diagnostics for the unrestricted market model indicate that instances of residual serial correlation are comparable to that for the benchmark model (Section 10.4.). Comparisons of the residual variance indicate that the mean residual variance of the unrestricted market model is still significantly higher than that of the benchmark model, although it is lower than that of the restricted model. The lower mean standard errors relative to those of the restricted model can explain the reductions in the understatement of the significance of factors (Section 10.3.1.; Section 10.5.). The residual variance structures of the unrestricted market model remain more complex relative to those of the benchmark model, as evident from a greater number of significant GARCH(*p,q*) models applied. Conditional variance structures are simpler relative to those underlying the restricted model. Interestingly, the inclusion of $IM\varepsilon_t$ in the unrestricted model results in a somewhat closer approximation of the conditional variance structures to those observed for the benchmark model (Section 10.5.).

Overall, the results suggest that the inclusion of $M\varepsilon_t$ in the unrestricted market model translates into improvements in predictive ability relative to the restricted model although predictive ability does not approximate that of the benchmark model. Mean errors for the unrestricted market model are still significantly larger than those of the benchmark model, although they are significantly lower relative to those of the restricted model. The mean Theil *U* statistic remains significantly higher than that of the benchmark model but is significantly lower than that of the restricted model. For the unrestricted model, the incorporation of the second residual market factor results in a mean bias proportion that approximates that of the benchmark model and is lower than that of the restricted model. Nevertheless, the mean variance proportion is still greater than that of the benchmark model. This is also the case for the covariance proportion that suggests that the proportion of systematic error attributable to the model structure, as opposed to the proportion of prediction errors attributable to random characteristics of the data, is higher than that of the benchmark model (Section 10.6.). These inferences, relating to the mean variance and covariance proportions, are true for both unrestricted specifications.

The analysis in Section 10.7. and Section 10.8. confirms that the unrestricted market model and the unrestricted model suffer from factor omission. Aside from a single instance, the null hypothesis that the factors that comprise the factor analytic augmentation are jointly insignificant is rejected for each sector, for both unrestricted specifications. In other words, there are other factors, reflected in f_{1t} and f_{2t} aside from the macroeconomic factors and the residual market factors, which impact returns. The influences of these factors are omitted from both unrestricted specifications and are not reflected by the residual market factors. Factor analysis of the resultant residual correlation matrices confirms this. For both specifications, a significant amount of shared variance associated with omitted common factors is reflected in the residuals (Section 10.7.). The greatest reduction in residual interdependence, relative to the restricted model, is attributable to the inclusion of $M\varepsilon_t$. The impact of $IM\varepsilon_t$ is almost unnoticeable when the residual correlation matrices derived from the unrestricted specifications are compared (Table 10.7.; Table 10.8.). The mean levels of residual correlation are still higher than those of the benchmark model and each residual correlation matrix exhibits a greater number of significant instances of residual correlation.

This is confirmed by the Jennrich (1970) test which shows that for both specifications, the residual correlation matrix structure does not approximate that of the benchmark model (Table 10.9.). This confirms that $M\varepsilon_t$ is not an adequate proxy for omitted factors and also that a two residual market factor approach does not alleviate factor omission bias.

In conclusion, both unrestricted models do not approximate the benchmark model in numerous aspects. The relative importance of factors remains understated, residual variance is inflated, the conditional variance structures are more complex than those of the benchmark model and the models underperform the benchmark model in terms of explanatory power and their ability to approximate the true return generating process. The validity of these specifications is challenged by the presence of common factors that can be extracted from the residuals, widespread significant pairwise residual correlation and the failure of the respective correlation matrices to approximate that of the benchmark model. Researchers and students of the APT should be mindful that the seemingly convenient use of a residual market factor or even multiple residual market factors to resolve underspecification in macroeconomic linear factor models may be ineffective. Failure to resolve underspecification will have adverse implications for the estimation and interpretation of the linear factor model – the building block of the APT – and tests of the APT relation.

Chapter 11 concludes this study, summarises the key aspects of the literature and the findings, suggests potential reasons for these findings and outlines avenues for further research and the limitations of this study.

CHAPTER 11

CONCLUSIONS AND RECOMMENDATIONS

11.1. INTRODUCTION

This chapter concludes this study. This study investigates the ability of the residual market factor approach to resolve underspecification in macroeconomic linear factor models that underpin the macroeconomic APT. Consideration is given to the impact of factor omission on inferences relating to factor significance, parameter bias, the robustness of a specification, residual variance, the structure of the conditional variance and predictive ability. The literature suggests that factor omission impacts each of these aspects (Section 5.3.1; Section 5.4.2.). This study also considers the structure of the resultant correlation matrices underlying each linear factor model specification, an aspect that is often not considered extensively. This latter consideration constitutes an investigation into the validity of the diagonality assumption underlying the linear factor model.

What follows in Section 11.2. is a summary of the literature and the theoretical basis of the study. Section 11.3. summarises the main findings and Section 11.4. outlines some possible reasons for the findings reported in this study. Section 11.5. proposes a factor analytic augmentation as a practicable and easily implementable solution that can be applied to mitigate factor omission in macroeconomic linear factor models. Limitations and areas for further research are outlined in Section 11.6. Section 11.7. concludes by emphasising the need for researchers and practitioners to be mindful of the consequences of underspecification when constructing and estimating macroeconomic linear factor models.

11.2. SUMMARY OF THE LITERATURE

Early empirical applications of the APT rely upon factor analytic techniques to identify the number of factors that feature in the return generating process and, subsequently, in the APT relation. The criticism directed at this approach spurs the development of the macroeconomic APT, following the work of Chan *et al.* (1985), Chen *et al.* (1986) and Hamao (1988).

In the macroeconomic APT, macroeconomic factors fulfil the role of pre-specified proxies for pervasive influences in stock returns that are represented by statistical factors in earlier studies. The macroeconomic APT is widely applied in the literature as a motivation and a theoretical basis for the study of the return generating process or asset pricing or in the consideration of both aspects simultaneously. Underpinning the literature is the macroeconomic linear factor model – an essential construct which must be specified and subsequently estimated (Section 2.4.).

Burmeister and Wall (1986) introduce the conventional residual market factor. Through orthogonalisation, the residual market factor permits a breakdown of total return variation attributable to extra-market factors and the residual market factor. The residual market factor is hypothesised to reflect the influence of omitted and unobservable factors that are not included in the linear factor model. The residual market factor can be constructed from a well-diversified portfolio, usually a broad domestic market aggregate (Section 3.2.). The literature has readily employed the residual market factor to proxy for omitted influences and to resolve underspecification in macroeconomic linear factor models. The literature also proposes the use of two residual market factors (Section 3.4.).

The role of global influences on stock returns is set out in Chapter 4. Markets exhibit interdependence that may be global and regional in nature and stock returns respond to macroeconomic news emanating from certain important markets, which can be termed information leaders (Section 4.2.1.; Section 4.2.2.). The APT provides for the role of global influences in the linear factor model by assuming either partial market integration/segmentation, full integration or complete segmentation. The latter two are seen as extremes and a more realistic assumption is that of partial integration/segmentation (Section 4.3.1.). This assumption allows a mixture of global and domestic factors to enter the linear factor model. Notably, a single factor that is assumed to proxy for global influences emerges from the literature - the MSCI World Market Index. This specific index is widely used in the literature to proxy for global macroeconomic influences and appears to outperform pre-specified global macroeconomic factors in explaining returns in the linear factor model (Section 4.3.2.; Section 4.4.). Because of the widespread use of this index in the literature and the practicability of including it as a factor in the linear factor model, this factor is chosen to derive the second residual market factor (Section 4.5.). The second residual market factor fulfils the role of a test factor in this study, which should be irrelevant if the conventional residual market factor accounts for all omitted influences.

Chapter 5 discusses underspecification and its impact on the APT. Underspecification of the linear factor model is likely. This is attributable to the unknown structure of the true return generating process, the unavailability of data, the omission of seemingly irrelevant factors, influences that are not quantified by macroeconomic factors, the principle of parsimony and the potentially incorrect assumption of linearity of the APT linear factor model (Section 5.2.). Underspecification is associated with intercept and coefficient bias, upward bias in the residual variance and coefficient standard errors, misleading inferences, unreliable and inaccurate model predictions and induces serial correlation and heteroscedasticity in the residuals, which can also impact inferences (Section 5.3.1.). The immediate assumptions underlying the linear factor model, that of residual diagonality and the absence of endogeneity, may also be violated. However and specifically, not much consideration is given to the validity of the diagonality assumption in the literature and specifically within the context of the macroeconomic linear factor model (Section 5.4.1.). If a linear factor model is underspecified, this will be reflected in the resultant residual correlation matrix, in the intercepts of the linear factor model and predictive ability, which will deteriorate. In the latter case, this is reflected by larger mean errors and other measures of predictive accuracy. The literature also suggests that the structure of the conditional variance will be impacted and that the linear factor model may be misidentified, as relevant factors will not be recognised as such. In the context of the APT relation, coefficient bias in the linear factor model may translate into incorrect inferences relating to the pricing of factors as factor betas will proxy for other factors that have been not been considered in the linear factor model. Perhaps most concerningly, underspecification may result in erroneous rejections of the validity of the APT (Section 5.4.2.). In Section 5.4.3., it is argued that because the linear factor is an underpinning construct of the APT, the consequences of underspecification on the linear factor model and the ability of the residual market factor to resolve underspecification should be investigated. These aspects have not been comprehensively considered in the literature.

11.3. SUMMARY OF THE FINDINGS

This study sets out to investigate whether the conventional residual market factor is a proxy for omitted factors and resolves underspecification. In doing so, the consequences of underspecification for the macroeconomic linear factor model are considered. Specifically, four research questions are posed in Section 1.2. and investigated.

The first research question relates to the ability of pre-specified macroeconomic factors to proxy for pervasive influences in returns. The return generating process is a complex construct. Evidence of this complexity is provided by the approach applied to identifying the factors that proxy for the pervasive influences in stock returns. A total of 52 macroeconomic factors are considered as candidate pre-specified proxies for the pervasive influences in stock returns (Table 6.4.). If considered in their contemporaneous form and up to three lag orders, then the candidate macroeconomic factor proxy set comprises a total of 208 permutations of factors. Out of these 208 permutations, only seven factors meet the required criteria set out for further consideration. These factors are BP_{t-1}, $LEAD_{t-1}$, BUS_t, USD_t, MET_t, LTY_t and TLI_t. While regressions of factor scores derived from returns on the industrial sectors onto the seven macroeconomic factors confirm that these macroeconomic factors are proxies for pervasive influences in returns, they also suggest that macroeconomic factors are poor proxies for pervasive influences. Even when the two residual market factors are incorporated into the factor regressions (Table 7.7.), the seven macroeconomic factors and the residual market factors yield a poor approximation of the factor scores. The difficulty in identifying macroeconomic factors that proxy for the pervasive influences in stock returns and the poor ability of these factors to approximate these pervasive influences is indicative of the limitations of the broader macroeconomic APT and the underpinning macroeconomic linear factor model. In conclusion, pre-specified macroeconomic factors are poor proxies for pervasive influences in stock returns.

The second research question relates to the impact of factor omission on the linear factor model within the context of the APT framework. This is investigated by juxtaposing three specifications against a benchmark model. The benchmark model combines the seven macroeconomic factors, the two residual market factors and factor scores derived from the residuals of a reduced form of the benchmark model that comprises the macroeconomic factors and the two residual market factors. The benchmark model appears to be well-specified. Of all specifications considered, this is the only specification for which mean errors are not significantly different from zero and there is no evidence of pervasive factors in the residuals of this specification (Table 8.4.; Table 8.5.). Furthermore, the resultant residual correlation matrix is void of widespread significant and systematic residual correlation (Table 8.7.). It appears that the macroeconomic factors, the two residual market factors and the factor analytic augmentation capture most of the pervasive influences in stock returns. However, the widespread significance of the statistical factors comprising the factor analytic augmentation suggests that there are other pervasive influences in stock returns that are not adequately proxied for by the macroeconomic and residual market factors in the first instance (Table 8.1.).

The specification that is estimated next is the restricted model, comprising the seven identified macroeconomic factors. This model performs poorly in numerous aspects relative to the benchmark model. The omission of the residual market factors and the factor analytic augmentation translates into a deterioration of explanatory power, a greater deviation from the true return generating process (Section 9.3.1.), inflated residual variance, larger mean errors, more complex conditional variance structures (Section 9.5.) and poor predictive ability (Section 9.6.). Arguably, the most severe consequence of underspecification is the misidentification of the macroeconomic factors that are systematic in nature. Such misidentification will translate into a misidentified linear factor model and by implication, a misidentified APT relation. The underspecification of the implied APT relation that will follow from a misidentified linear factor model may lead to erroneous rejections of the validity of the APT and erroneous interpretations of priced factors. Priced factors may reflect the influences of other factors and idiosyncratic factors (analogously industry-specific factors) may appear to be priced. Pricing, however, may be attributable to the failure to account for pervasive influences in the linear factor model and not the failure of the theoretical foundations of the APT framework (Section 2.3.1.; Section 5.4.2.). Moreover, the diagonality assumption is widely violated. The residual correlation matrix is characterised by widespread pairwise residual correlation and does not differ in structure to that of the actual return series (Section 9.7.; Section 9.8.). This is a serious indictment on the ability of macroeconomic factors to account for co-movement in returns attributable to underlying pervasive factors. Researchers should be careful when interpreting the results of a linear factor model that relies only on macroeconomic factors. Researchers should also be aware that the results of an APT relation that follows from such a linear factor model should be interpreted with caution. In conclusion, factor omission adversely impacts numerous aspects of the linear factor model and can affect the interpretation of the APT relation.

The third research question considers the efficacy of using a conventional residual market factor to resolve underspecification. The unrestricted market specification, which incorporates the macroeconomic factors and the conventional residual market factor, is estimated to determine whether the inclusion of a residual market factor derived from returns on the JSE All Share Market Index resolves underspecification. The premise is that if the conventional residual market factor resolves underspecification, then the various aspects of the unrestricted market model will be comparable to that of the benchmark model. The unrestricted specification comprises macroeconomic factors and incorporates a second residual market factor in addition to the conventional residual market factor, derived from returns on the MSCI World Market Index (Section 4.3.1.; Section 4.3.2.). The premise underlying the unrestricted market model is that a second residual market factor should be redundant if the conventional residual market factor is an adequate proxy for omitted factors.

The analysis in Chapter 10 suggests that the inclusion of a conventional residual market factor, $M\varepsilon_t$, improves estimation results relative to the restricted model. All factors are now identified as systematic in the unrestricted market model; the model exhibits higher explanatory power and better approximates the true return generating process (Table 10.1.). The same can be said about the model diagnostics. Notably, the number of significant instances of residual serial correlation is now comparable to that of the benchmark model (Section 10.4.). However, the unrestricted market model underperforms the benchmark model in numerous aspects. The importance of certain factors is still understated, the magnitude of the residual variance exceeds that of the benchmark model for individual sectors, coefficient standard errors are inflated and the conditional variance structures underlying the individual industrial sectors are more complex relative to those of the benchmark model (Table 10.3.) Mean errors are significantly greater than those of the benchmark model and predictions show significantly greater bias relative to the benchmark model (Table 10.4.). The LR test confirms that the unrestricted market model is underspecified and that factors have been omitted. Factor analysis indicates that factors that can explain a significant proportion of shared variance can be extracted from the residuals of the unrestricted market model (Table 10.5.). Finally, an analysis of the residual correlation matrix indicates that although there is a general decrease in the magnitude of residual correlation and there are fewer instances of significant pairwise residual correlation relative to the restricted model, the residual correlation matrix is not comparable to that derived from the benchmark model (Table 10.6; Table 10.7.). Much of the co-movement attributable to omitted pervasive factors is reflected in the residual correlation matrix which exhibits widespread significant violations of the diagonality assumption. In conclusion, the conventional residual market factor does not appear to be an adequate proxy for omitted factors nor does it adequately resolve underspecification.

The fourth question relates to the efficacy of using a second residual market factor to resolve any remaining underspecification. This factor is treated as both a test factor and a proxy for any remaining factors, motivated by the importance of global influences in contemporary financial markets. This second residual market factor, $IM\varepsilon_t$, is derived from returns on the MSCI World Market Index, an index that is widely used in the literature as a proxy for global influences in stock markets. The most significant finding is that $IM\varepsilon_t$ is statistically significant in the unrestricted model for more than half of the industrial sectors in the sample (Table 10.1.). This suggests that the conventional residual market factor fails to account for relevant global influences. This constitutes further evidence against the efficacy of $M\varepsilon_t$ as a proxy for omitted influences. Aside from this finding, the contribution of the second residual market factor to improving explanatory power, to approximating the return generating process and to improving predictive ability is negligible (Section 10.3.1.; Section 10.6.). The only aspects that are significantly impacted are residual variance and the conditional variance structures. The number of sectors for which residual variance is comparable to that of the benchmark model in magnitude is greater relative to that of the unrestricted market model and the conditional variance structures more closely approximate those of the benchmark model (Table 10.3.; Section 10.5.). Nevertheless, the LR test for omitted factors and factor analysis on the resultant residuals confirm that the unrestricted model remains underspecified and that there are pervasive factors that are relegated to the residuals (Table 10.5.). The mean level of residual correlation is not much lower than that of the unrestricted market model and the residual correlation matrix continues to reflect mostly positive and significant interdependence (Table 10.6.; Table 10.8.). In conclusion, the unrestricted model, much like the unrestricted market model, fails to approximate the numerous aspects of the benchmark model. Therefore, a two residual market factor approach, although easily implementable and practicable, fails to adequately resolve underspecification. In conclusion, a second residual market factor derived from a widely used international market index is not effective in resolving remaining underspecification and suggests that a conventional residual market factor does not reflect all influences.

11.4. A COMMENT ON THE ADEQUACY OF THE RESIDUAL MARKET FACTOR

Although the investigation of the reasons for the failure of macroeconomic factors and the residual market factors to adequately proxy for the pervasive influences in returns is beyond the scope of the study, it is worth noting that some potential reasons have been suggested by the literature.

Born and Moser (1988: 289) state that the formation of the true market portfolio is an aggregation process that will reflect the influences of pervasive factors. However, the true market portfolio is unobservable (Brown & Brown, 1987: 26). This implies that any market proxy will fail to reflect all relevant influences. Consequently, the residual market factor will never be an adequate proxy for all omitted factors. It can be argued that in this study, the respective indices used to derive the two residual market factors, namely the JSE All Share Index and MSCI World Market Index, are imperfect proxies for the true market portfolio and do not reflect all relevant influences. It follows that a solution would be to specify a broader market index and then to derive a residual market factor from this index. This is a seemingly appealing solution but is characterised by two limitations. Firstly, one could formulate increasingly broader indices and use these to derive residual market factors that would ideally reflect all relevant influences to the sample at hand. However, this however detracts from the ease of applying the residual market factor approach to resolving underspecification. One would need to specify, test and then respecify a market index to determine whether it reflects all relevant influences – a potentially lengthy, if not futile, exercise. Secondly, Brown and Brown (1987: 31) argue that the quest for an all-inclusive market portfolio is a futile one. This implies that although a market benchmark could be specified and then respecified, a representation of the true market portfolio will never be achieved in the first place (Section 11.6.). Consequently, a residual market factor derived from a market proxy will never be a proxy for all omitted factors.

Another potential reason, suggested by the work of Van Rensburg (2000), relates to the structure of a given stock market and the general structure of the economy. The author uses two residual market factors to capture the dichotomy of the return generating process underlying South African mining and industrial stocks. Following from this, it may be that a given economy is driven by two or more large industrial sectors. However, a general domestic market index may be representative of all economic sectors equally and may underweight the economic sectors that drive the economy. It can therefore be argued that such a proxy will fail to reflect all relevant influences. The solution would be to carefully consider the market and economic structure and to use appropriate industrial sector indices to proxy for relevant pervasive influences.

Next, and for numerous reasons, a residual market factor derived from a broad equity index may not be an optimal proxy for omitted factors. As suggested in Brown *et al.* (2009: 302), a residual market factor derived from a bond market index may provide a better proxy for omitted factors. The authors report that the $\bar{R}^2$ for a regression of returns on the Citigroup World Government Bond Index onto the set of global macroeconomic factors in their study is more than double that of a regression of these factors onto returns on the MSCI World Market Index (equity). This suggests that the correlation between returns on a bond market index and the global macroeconomic environment is stronger and potentially more widespread in terms of the macroeconomic influences that are reflected in such a factor. A residual market factor derived from a bond index may be more reflective of prevailing economic conditions than one derived from a purely equity-based index (see Selection 11.6).

McQueen and Roley (1993: 694) show that the impact of macroeconomic factors on stock markets is conditional upon the state of the economy (Section 3.5.). For example, the authors find that during a “high” state, returns on the S&P 500 index respond to announcements in five of the eight economic factors considered, whereas during “low” states, the S&P500 is unimpacted by news relating to these macroeconomic factors. Therefore, markets may be more responsive to macroeconomic information during certain economic states and less so during other economic states. Moreover, this suggests that macroeconomic linear factor models may be underspecified not because of the omission of factors, but because factors that proxy for pervasive influences during certain parts of the sample period do not proxy for pervasive influences during other parts of the sample period. Similarly, it can be argued that because of the changing state of the economy, the residual market factor is a better proxy for relevant influences during certain economic states than others. That markets may respond differently to the macroeconomic environment is suggested by Lo’s (2004) adaptive market hypothesis. Lo (2004: 24) suggests that one of the implications of this hypothesis is that the relationship between risk and reward is unlikely to be stable over time as such a relation is determined by market preferences and institutional aspects, which themselves change over time. In the present context, the implication is that the ability of macroeconomic factors to explain returns and the efficacy of a residual market factor to proxy for omitted influences will change over time, potentially rendering a macroeconomic linear factor model unable to adequately characterise the return generating process.

Following from the adaptive market hypothesis is the controversy regarding the validity of the efficient market hypothesis. Chan *et al.* (1985: 452) argue that in an efficient market, macroeconomic information should rapidly be reflected in a market aggregate. By extension, this also relates to the residual market factor. This premise is what affords the residual market factor the role of a proxy for information and omitted factors. If markets are not efficient, then it is likely that macroeconomic information is not reflected in returns accurately or on time. A macroeconomic linear factor model will not adequately characterise the return generating process and the residual market factor will not be a proxy for omitted factors. Lo (2004: 17; 20) suggests that this question is as yet unanswered, stating that "economists have not yet reached a consensus about whether markets – particularly financial markets – are in fact efficient." It is against a behavioural backdrop that the author goes onto motivate for the adaptive markets hypothesis.

Finally, Baker and Wurgler (2007: 129) argue that an assumption underlying standard models in finance, such as the CAPM and the APT, is that investors are rational. Consequently, investors force prices to be equal to the rational present value of future expected cash flows. Within the context of the macroeconomic APT, this suggests that investors react rationally to macroeconomic news, which is a determinant of stock prices. The APT does however provide for the role of sentiment in determining returns. Spronk and Hallerbach (1997: 123) suggest that a residual market factor may also reflect behavioural factors. Nevertheless, Baker and Wurgler (2007) argue that numerous stock market, events such as the Great Crash of 1929, the Black Monday crash of 1987 and the Internet Bubble towards the end of the 1990s and the beginning of the 2000s, defy explanation. This suggests that investors react to news irrationally at times. It follows that a model that assumes rationality and is constructed on this premise may not yield an adequate representation of the return generating process. Furthermore, it can be argued that investors are both rational and also respond to sentiment. As a result, macroeconomic linear factor models may not be sufficiently reflective of sentiment. Although sentiment will be reflected in the residual market factor indirectly, as suggested by Spronk and Hallerbach (2017), sentiment may be far more important than implicitly provided for by the APT. This calls for an explicit incorporation of a sentiment measure or measures into constructs of the linear factor model. A specification that is constructed on the premise of rationality but also explicitly takes into account behavioural aspects may provide a better approximation of the true return generating process relative to a linear factor model that comprises only macroeconomic factors and the residual market factor.

The present discussion demonstrates that there is a plethora of reasons why macroeconomic linear factor model may be underspecified and why the use of a residual market factor is unlikely to resolve underspecification.

11.5. FACTOR ANALYTIC AUGMENTATION AS A SOLUTION

The macroeconomic linear factor model and the residual market factor are easily understood, elegant and appealing constructs. Yet, this study suggests that a combination of macroeconomic factors and the residual market factor in a linear factor model will not produce an adequate representation of the return generating process. The pertinent question is whether the macroeconomic linear factor model, and by extension, the macroeconomic APT, can be salvaged. It is argued here that these constructs can be salvaged and that the solution lies in the factor analytic augmentation. The factor analytic augmentation presents a relatively simple solution to the underspecification problem.

The linear factor model can be estimated within two contexts. In the first instance, the linear factor model is estimated to derive inputs for the APT relation in the form of the coefficients associated with the macroeconomic factors that feature in the linear factor model. This requires that the same specification is estimated across a number of assets, be they individual stocks or portfolios (Section 2.3.4.; Section 2.4.). This approach is followed in this study and in APT literature in general. In the second instance, the macroeconomic linear factor model is estimated to study the behaviour of a single series and the estimated factor sensitivities are not used for any other purposes than direct interpretation and inference making about the structure of the return generating process. Examples of such studies are those of Sadorsky and Henriques (2001), who investigate the behaviour of returns for the Canadian paper and forest products industry, Sadorsky (2001), who investigates the factors driving returns on an index comprising Canadian oil and gas companies and Szczygielski and Chipeta (2015), who model the return generating process underlying the South African stock market (Section 2.4.).

In the first instance, a similar approach to that followed in this study can be applied. In the first step, a reduced form version of the benchmark model comprising the desired macroeconomic factors and a residual market factor is estimated (Section 6.4.1.; Section 8.2.; equation (8.1)). In the second step, the resultant correlation matrix is factor analysed and any extracted factors are appended to the initial specification to produce an expanded form specification, an equivalent of the benchmark model in this study (equation (8.2)). These statistical factors are then hypothesised to represent any influences, whether purely macroeconomic or behavioural in nature, that have not been explicitly reflected by the macroeconomic factor set and the residual market factor in a proposed formulation of the macroeconomic linear factor model. As demonstrated in Chapter 8, this approach produces relatively desirable results.

In the second instance, the linear factor model is estimated for a single return series and is of direct interest as a representation of the return generating process for that series. In the first step, a broader set of return series from the same market is factor analysed to extract factor scores representative of pervasive influences. In the second step, the extracted factor scores are orthogonalised against the macroeconomic factor set that is used to represent the proposed linear factor model, and, if included, the residual market factor. In the third step, the orthogonalised factor scores are used to augment the specification of interest. The appeal of this approach is that while it requires return data beyond that of the series of direct interest, the linear factor model is still tailored to the series of interest. In constructing the model, only factors that are of interest to the subject series need to be considered, unlike in the APT, which requires that the macroeconomic factors selected have a systematic impact across the sample. The orthogonalised factor scores now represent other pervasive factors that prevail in the market of interest and will go some way to resolving any underspecification that may arise.

It is hoped that this approach aides in resolving underspecification while permitting the continued use of pre-specified and interpretable macroeconomic factors together with the residual market factor.

11.6. LIMITATIONS AND AREAS FOR FURTHER RESEARCH

Admittedly, this study suffers from limitations although these limitations present avenues for further research.

The first limitation lies in that only a single stock market is considered in this study, the South African stock market. Inferences regarding the adequacy of the macroeconomic linear factor model and the residual market factor may not hold for other markets. It may be that macroeconomic factors are effective proxies for the pervasive influences in returns in other markets. However, the findings in Connor (1995) for the US market make this assertion questionable. Also, it may be that in other markets, the residual market factor is an adequate proxy for omitted factors. Therefore, it is suggested that the ability of macroeconomic factors to proxy for pervasive influences in stock returns and the efficacy of the residual market factor are considered in a similar manner for other markets. This could be done by following the approach set out in Section 6.3.2., namely that of relating derived factor scores to macroeconomic factors and a residual market factor or residual market factors.

The second limitation is that this study relies on two widely used market indices to derive the residual market factors. As stated in Section 11.4., it is possible that these indices are not broad enough to aggregate the influence of all relevant return generating factors and may be poor proxies for the market portfolio in the first place. For example, the MSCI World Market Index comprises large and mid-cap stocks across 23 developed markets. The index excludes emerging markets (MSCI, 2018). A market index that represents the capitalisation on both developed and emerging markets will be broader, and as argued by Born and Moser (1988), is likely to be a better aggregate of the influences driving returns. Also, Wei (1988: 889) recognises that a poorly diversified market proxy may lead to a rejection of the null hypothesis that the residuals for a given linear factor model are zero. This suggests that such a market proxy fails to account for all influences not reflected by the macroeconomic factor set. Brown and Brown (1987: 31) find that a successively broader market proxy matters for evaluating the historical performance on investment portfolios and argue that the market should be defined in terms of the relevant asset universe. It follows that a further avenue of research is to consider how the composition of the market proxy will impact the residual correlation matrix, within the context of the macroeconomic linear factor model (as opposed to performance evaluation). It may be worthwhile establishing whether there is a point at which broadening the market proxy has no impact on results and yields no reductions in pairwise residual interdependence. At this point, any remaining underspecification may no longer impacts results. Additionally, motivated by the work Brown *et al.* (2009), consideration could also be given to the ability of a bond market portfolio to proxy for omitted and unobserved factors in linear factor model specifications of the return generating process. It may be that a non-equity aggregate is a better proxy for omitted factors. Such an investigation could compare the results of a linear factor model that combines macroeconomic factors and a residual market factor derived from an equity market index to a linear factor model that combines macroeconomic factors and a residual market factor derived from a bond market aggregate.

A third limitation of this study is that it investigates the consequences of underspecification in the linear factor model and does not explicitly extend this investigation to the APT relation. Although the study is contextualised within the APT and inferences are made with regard to the potential impact on the APT relation, such as a possible rejection of the validity of the

APT and the erroneous pricing of factors attributable to coefficient bias, the impact of factor omission in the underpinning linear factor model is not empirically related to the APT relation in this study. This poses a further avenue for research that could take a similar approach as that undertaken in this study. A benchmark linear factor model that incorporates a factor analytic augmentation, a restricted model and unrestricted models that include residual market factors could be estimated and the impact on the validity of the APT relation and the pricing implications of factor omission could be investigated. Results of the APT relation underpinned by a restricted specification could be compared to those of the APT relation based upon the benchmark model. The existence of confounding relationships - those that arise because of correlation between factors but are not directly attributable to a certain factor that is included in a specification - between expected returns and sensitivities to specific factors could be identified and investigated (Jorion 1991: 366; Section 5.4.2.). Furthermore, research that investigates the impact of an underspecified linear factor model on the APT relation could also provide insight into how factor omission impacts classic tests of the APT that rely upon idiosyncratic factors.

A fourth limitation is that this study only superficially investigates the impact of factor omission on conditional variance. The results are mixed. In contrast to the work of Bera *et al.* (1988), who argue that factor omission impacts conditional heteroscedasticity, this study does not find conclusive evidence that this is the case. However, it finds that factor omission impacts the conditional variance structure. Linear factor models that omit factors appear to have more complex conditional variance structures. This study also confines itself to using two ARCH/GARCH specifications to describe the conditional variance, the ARCH(p) and GARCH(p,q) models. The impact of underspecification on the structure of the conditional variance may be investigated further by using a broader set of ARCH/GARCH-type models. In doing so, it should be established why conditional heteroscedasticity does not appear to be impacted by factor omission as reported in this study. It may be that this is because different conditional variance specifications are estimated as opposed to a single model, the ARCH(p) model as in Bera *et al.* (1988). Consequently, it may be that the impact of underspecification is primarily reflected in the overall conditional variance structure and not conditional heteroscedasticity, as suggested by this study. This may be confirmed through further dedicated empirical work. Additionally, there are relatively few studies that consider the impact of the conditional variance structure on beta estimates in the conditional mean. Aside from Bera *et al.* (1988), the author is aware of only Hamilton (2010), Brzeszczyński *et al.* (2011) and Armitage and Brzeszczyński (2011), who suggest that the structure of conditional variance impacts coefficient estimates and that there may be value in permitting the structure of conditional variance to impact betas estimates. It follows that a dedicated study of the impact of the structure of conditional variance on beta estimates and a comparison of the benefits of such beta estimates with those derived using least squares may be of value.

Finally, a rather “purist” approach of assuming a strict factor structure is undertaken in this study. This is the assumption that underlies the Fama-Macbeth two-step methodology used to estimate factor coefficients for use in the APT relation (Section 2.2.). However, as suggested by Clare *et al.* (1997b: 652), methodologies such as NL3SLS, permit an approximate factor structure that allows for pairwise residual correlation and leads to efficiency gains, improving pricing results. This study assumes that a strict factor structure underlies the linear factor model and consequently applies an econometric methodology that does not explicitly permit for an approximate factor structure. It can nevertheless be argued that this does not detract from the findings of this study. Within the context of the APT, techniques that permit for a violation of the diagonality assumption are appropriate. However, the linear factor model is not always of interest as an underpinning construct of the APT relation which requires an assumption about the validity of the diagonality assumption. In other words, the linear factor model need not be estimated as a system of regressions for numerous series but may be estimated for a single series or market to investigate the structure of the return generating process. In this case, the question of the structure of the residual correlation matrix falls away. Yet, the challenge of underspecification and the associated consequences remains. Furthermore, ignoring pairwise residual correlation equates to ignoring information that is reflected in the residuals of the linear factor model and presents a failure to interrogate the reasons for the violation of the diagonality assumption.

11.7. CONCLUSION

The motivation for this study is the importance of the construct underpinning the APT relation - the linear factor model. It is hoped that this study provides a comprehensive insight into the consequences of factor omission on the linear factor model and the inferences that arise from the estimation of this construct.

The main findings are that the use of macroeconomic factors may not translate into a well-specified model and that the residual market factor fails to resolve factor omission bias. Macroeconomic factors are poor proxies for the pervasive factors in returns and the interpretation of an underspecified linear factor model will pose challenges and questionable inferences may be arrived at. The widespread approach of using a residual market factor to resolve factor omission will not resolve underspecification; and a two residual market factor approach that relies on widely used market indices is not effective either. The inclusion of a factor analytic augmentation may assist in deriving a more reliable and interpretable linear factor model that incorporates pre-specified macroeconomic factors.

There is value in constructing and estimating macroeconomic linear factor models. Not only do such models provide insight into what macroeconomic forces drive financial markets, they also underpin the APT framework. The recommendation to researchers and practitioners who study the impact of the macroeconomic environment on financial markets or seek to study pricing implications by applying the APT relation is that an inherently adequately specified macroeconomic linear factor model should not be assumed in the first instance. Any proposed specification that combines macroeconomic factors with a residual market factor should be interrogated further to determine whether any systematic factors have been omitted.

LIST OF REFERENCES

Abdi, H. (2003). Factor Rotations in Factor Analyses (sic). In M. Lewis-Beck, A, Bryman & T. Futing (Eds.). *Encyclopedia of Social Sciences Research Methods.* Thousand Oaks, CA: Sage.

Abugri, B. A. (2008). Empirical relationship between macroeconomic volatility and stock returns: Evidence from Latin American markets. *International Review of Financial Analysis*, *17*(2), 396-410.

Adams, C., & Coe, D. (1990). A systems approach to estimating the natural rate of unemployment and potential output for the United States. *Staff Papers (International Monetary Fund), 37*(2), 232-293.

Aho, K., Derryberry, D., & Peterson, T. (2014). Model selection for ecologists: The worldviews of AIC and BIC. *Ecology*, *95*(3), 631-636.

Akgiray, V. (1989). Conditional heteroscedasticity in time series of stock returns: Evidence and forecasts. *Journal of Business, 62*(1), 55-88.

Amenc, N., & Le Sourd, V. (2003). *Portfolio theory and performance analysis.* Chichester: John Wiley & Sons Ltd.

Andersen, R. (2008). *Quantitative applications in the social sciences: Modern methods for robust regression.* Thousand Oaks, CA: SAGE Publications Ltd

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Vega, C. (2003). Micro effects of macro announcements: Real-time price discovery in foreign exchange. *American Economic Review*, *93*(1), 38-62.

Antin, D. (2013). The South African mining sector: An industry at a crossroads. *Economy Report South Africa*. Retrieved December 4, 2017 from the Hans Seidel Foundation website at: https://southafrica.hss.de/fileadmin/user_upload/Projects_HSS/South_Africa/170911_Migration/Mining_Report_Final_Dec_2013.pdf

Antoniou, A., Garrett, I., & Priestley, R. (1998). Macroeconomic variables as common pervasive risk factors and the empirical content of the arbitrage pricing theory. *Journal of Empirical Finance*, *5*(3), 221-240.

Aretz, K., Bartram, S. M., & Pope, P. F. (2010). Macroeconomic risks and characteristic-based factor models. *Journal of Banking and Finance*, *34*(6), 1383-1399.

Armitage, S., & Brzeszczyński, J. (2011). Heteroscedasticity and interval effects in estimating beta: UK evidence. *Applied Financial Economics*, *21*(20), 1525-1538.

Arshanapalli, B., & Doukas, J. (1993). International stock market linkages: Evidence from the pre-and post-October 1987 period. *Journal of Banking and Finance*, *17*(1), 193-208.

Artiach, T., Lee, D., Nelson, D., & Walker, J. (2010). The determinants of corporate sustainability performance. *Accounting & Finance*, *50*(1), 31-51.

Asteriou, D., & Hall, S.G. (2016). *Applied econometrics* (3rd ed.). London: Palgrave

Awartani, B. M., & Corradi, V. (2005). Predicting the volatility of the S&P-500 stock index via GARCH models: the role of asymmetries. *International Journal of Forecasting*, *21*(1), 167-183.

Azeez, A. A., & Yonezawa, Y. (2006). Macroeconomic factors and the empirical content of the arbitrage pricing theory in the Japanese stock market. *Japan and the World Economy*, *18*(4), 568-591.

Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, *66*(1), 47-78.

Baier, S. L., & Bergstrand, J. H. (2001). The growth of world trade: tariffs, transport costs, and income similarity. *Journal of International Economics*, *53*(1), 1-27.

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, *21*(2), 129-152.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, *9*(1), 3-18.

Barr, G.D.I. (1990). Macroeconomic identification of the pricing factors on the Johannesburg Stock Exchange. *South African Journal of Business Managemen*t, *21*(1), 17-16.

Barreto, H., & Howland, F.M. (2006). *Introductory econometrics: Using monte carlo simulation with Excel.* Cambridge: Cambridge University Press.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society A, 160*(901), 268-282.

Basu, S. (1977). Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *Journal of Finance*, *32*(3),663-682.

Basu, D., & Chawla, D. (2012). An empirical test of the arbitrage pricing theory—The case of Indian stock market. *Global Business Review*, *13*(3), 421-432.

Beenstock, M., & Chan, K. F. (1986). Testing the arbitrage pricing theory in the United Kingdom. *Oxford Bulletin of Economics and Statistics*, *48*(2), 121-141.

Beenstock, M., & Chan, K. F. (1988). Economic forces in the London stock market. *Oxford Bulletin of Economics and Statistics*, *50*(1), 27-39.

Beirne, J., Caporale, G. M., Schulze-Ghattas, M., & Spagnolo, N. (2010). Global and regional spillovers in emerging stock markets: A multivariate GARCH-in-mean analysis. *Emerging Markets Review*, *11*(3), 250-260.

Bera, A., Bubnys, E., & Park, H. (1988). Conditional heteroscedasticity in the market model and efficient estimates of betas. *Financial Review*, *23*(2), 201-214.

Berry, M. A., Burmeister, E., & McElroy, M. B. (1988). Sorting out risks using known APT factors. *Financial Analysts Journal*, *44*(2), 29-42.

Bilson, C. M., Brailsford, T. J., & Hooper, V. J. (2001). Selecting macroeconomic variables as explanatory factors of emerging stock market returns. *Pacific-Basin Finance Journal*, *9*(4), 401-426.

Birz, G., & Lott Jr, J. R. (2011). The effect of macroeconomic news on stock returns: New evidence from newspaper coverage. *Journal of Banking and Finance*, *35*(11), 2791-2800.

Bishara, A. J., & Hittner, J. B. (2012). Testing the significance of a correlation with nonnormal data: comparison of Pearson, Spearman, transformation, and resampling approaches. *Psychological Methods*, *17*(3), 399-417.

Blackwell, M. (2008). *Multiple hypothesis testing: The F-test.* Working Paper. Retrieved December 4, 2017, from the Matt Blackwell Research website at: http://www.mattblackwell.org/files/teaching/ftests.pdf

Blanchard, O. J. (1987). [Vector Autoregressions and Reality]: Comment. *Journal of Business and Economic Statistics*, *5*(4), 449-451.

Bliemel, F. (1973). Theil's forecast accuracy coefficient: A clarification. *Journal of Marketing Research*, *10*(4), 444-446.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307-327.

Bollerslev, T., Chou, R. Y., & Kroner, K. F. (1992). ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics*, *52*(1-2), 5-59.

Bollerslev, T., & Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, *11*(2), 143-172.

Bonett, D. G., & Wright, T. A. (2000). Sample size requirements for estimating Pearson, Kendall and Spearman correlations. *Psychometrika*, *65*(1), 23-28.

Born, J. A., & Moser, J. T. (1988). An Investigation into the role of the market portfolio in the arbitrage pricing theory. *Financial Review*, *23*(3), 287-299.

Bower, D. H., Bower, R. S., & Logue, D. E. (1984). Arbitrage Pricing Theory and utility stock returns. *Journal of Finance*, *39*(4), 1041-1054.

Bradfield, D. J. (1990). The influence of the NYSE on the risk of JSE stocks. *De Ratione*, *4*(1), 2-6.

Brauer, J., & Gómez-Sorzano, A. G. (2004). Homicide cycles in Colombia, 1950-1999. *International Journal of Applied Econometrics and Quantitative Studies*, *1*(1), 1-22.

Brennan, M. J., Chordia, T., & Subrahmanyam, A. (1998). Alternative factor specifications, security characteristics, and the cross-section of expected stock returns1. *Journal of Financial Economics*, *49*(3), 345-373.

Breusch, T. S. (1978). Testing for autocorrelation in dynamic linear models. *Australian Economic Papers*, *17*(31), 334-355.

Broadman, H. G. (2008). China and India go to Africa: New deals in the developing world. *Foreign Affairs*, *87*(2), 95-109.

Brooks, C. (2008). *Introductory econometrics for finance* (2nd ed.). New York: Cambridge University Press.

Brooks, C., & Tsolacos, S. (2010). *Real estate modelling and forecasting*. Cambridge: Cambridge University Press

Brown, K. C., & Brown, G. D. (1987). Does the composition of the market portfolio really matter? *Journal of Portfolio Management*, *13*(2), 26-32.

Brown, M. B., & Forsythe, A. B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association*, *69*(346), 364-367.

Brown, S. J., Hiraki, T., Arakawa, K., & Ohno, S. (2009). Risk premia in international equity markets revisited. *Pacific-Basin Finance Journal*, *17*(3), 295-318.

Brown, S. J., & Weinstein, M. I. (1983). A new approach to testing asset pricing models: The bilinear paradigm. *Journal of Finance*, *38*(3), 711-743.

Brzeszczyński, J., Gajdka, J., & Schabek, T. (2011). The role of stock size and trading intensity in the magnitude of the" interval effect" in beta estimation: Empirical evidence from the Polish capital market. *Emerging Markets Finance and Trade*, *47*(1), 28-49.

Bucevska, V. (2011). Heteroscedasticity. In M. Lovric (Ed.). *International Encyclopedia of Statistical Science.* 630-633. Berlin, Heidelberg: Springer.

Bureau for Economic Research (BER). (2015). *Purchasing Managers' Index (PMI): A monthly index of business conditions in the manufacturing sector.* Retrieved December 4, 2017 from the BER website at: https://www.ber.ac.za/knowledge/pkdownloaddocument.aspx?docid=6547

Bürki, A., & Gaskell, M. G. (2012). Lexical representation of schwa words: Two mackerels, but only one salami. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(3), 617-631.

Burmeister, E., & McElroy, M. B. (1991). The residual market factor, the APT, and mean-variance efficiency. *Review of Quantitative Finance and Accounting*, *1*(1), 27-49.

Burmeister, E., Roll, R., & Ross, S. (1994). *A practitioner's guide to arbitrage pricing theory.* In A practitioner's guide to factor models. Charlottesville: The Research Foundation of the Institute of Chartered Financial Analysts.

Burmeister, E., & Wall, K. D. (1986). The arbitrage pricing theory and macroeconomic factor measures. *Financial Review*, *21*(1), 1-20.

Campbell, J.Y., Lo, A.W., & MacKinlay, A.C. (1997). *The econometrics of financial markets.* Princeton, NJ: Princeton University Press.

Cauchie, S., Hoesli, M., & Isakov, D. (2004). The determinants of stock returns in a small open economy. *International Review of Economics and Finance*, *13*(2), 167-185.

Chan, K. C., Chen, N. F., & Hsieh, D. A. (1985). An exploratory investigation of the firm size effect. *Journal of Financial Economics*, *14*(3), 451-471.

Chan, K. C., Hendershott, P. H., & Sanders, A. B. (1990). Risk and return on real estate: Evidence from equity REITs. *Real Estate Economics*, *18*(4), 431-452.

Chang, S. J. (1991). A study of empirical return generating models: A market model, a multifactor model, and a unified model. *Journal of Business Finance & Accounting*, *18*(3), 377-391.

Charles, A. (2010). The day-of-the-week effects on the volatility: The role of the asymmetry. *European Journal of Operational Research*, *202*(1), 143-152.

Chen, M. H. (2010). Understanding world metals prices: Returns, volatility and diversification. *Resources Policy*, *35*(3), 127-140.

Chen, N. F. (1983). Some empirical tests of the theory of arbitrage pricing. *Journal of Finance*, *38*(5), 1393-1414.

Chen, N. F., Roll, R., & Ross, S. A. (1986). Economic forces and the stock market. *Journal of Business*, *59*(3), 383-403.

Chen, S. J., Hsieh, C. H., & Jordan, B. D. (1997). Real estate and the arbitrage pricing theory: Macrovariables vs. derived factors. *Real Estate Economics*, *25*(3), 506-523.

Chen, S. J., Hsieh, C., Vines, T., & Chiou, S. N. (1998). Macroeconomic variables, firm-specific variables and returns to REITs. *Journal of Real Estate Research*, *16*(3), 269-278.

Chen, S. J., & Jordan, B. D. (1993). Some empirical tests in the arbitrage pricing theory: Macro variables vs. derived factors. *Journal of Banking and Finance*, *17*(1), 65-89.

Chenhall, R. H., & Moers, F. (2007). The issue of endogeneity within theory-based, quantitative management accounting research. *European Accounting Review*, *16*(1), 173-196.

Chesher, A., & Jewitt, I. (1987). The bias of a heteroskedasticity consistent covariance matrix estimator. *Econometrica*, *55*(5), 1217-1222.

Chimanga, A., & Kotze, D. (2009). A multivariate analysis of factors affecting stock returns on the JSE. *African Finance Journal*, *11*(2), 80-96.

Cho, D. C. (1984). On testing the arbitrage pricing theory: Inter-battery factor analysis. *Journal of Finance*, *39*(5), 1485-1502.

Cho, D. C., Eun, C. S., & Senbet, L. W. (1986). International arbitrage pricing theory: An empirical investigation. *Journal of Finance*, *41*(2), 313-329.

Choi, J. J., & Rajan, M. (1997). A joint test of market segmentation and exchange risk factor in international capital market. *Journal of International Business Studies*, *28*(1), 29-49.

Claar, V. V. (2006). Is the NAIRU more useful in forecasting inflation than the natural rate of unemployment? *Applied Economics*, *38*(18), 2179-2189.

Clare, A. D., & Priestley, R. (1998). Risk factors in the Malaysian stock market. *Pacific-Basin Finance Journal*, *6*(1-2), 103-114.

Clare, A. D., & Thomas, S. H. (1994). Macroeconomic Factors, the APT and the UK stock market. *Journal of Business Finance and Accounting*, *21*(3), 309-330.

Clare, A., Priestley, R., & Thomas, S. (1997a). Is beta dead? The role of alternative estimation methods. *Applied Economics Letters*, *4*(9), 559-562.

Clare, A., Priestley, R., & Thomas, S. (1997b). The robustness of the APT to alternative estimators. *Journal of Business Finance and Accounting*, *24*(5), 645-655.

Clarke, K. A. (2005). The phantom menace: Omitted variable bias in econometric research. *Conflict Management and Peace Science*, *22*(4), 341-352.

Clements, M. P., & Galvão, A. B. (2008). Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States. *Journal of Business and Economic Statistics*, *26*(4), 546-554.

Collins, D., & Biekpe, N. (2003). Contagion and interdependence in African stock markets. *South African Journal of Economics*, *71*(1), 181-194.

Connor, G. (1995). The three types of factor models: A comparison of their explanatory power. *Financial Analysts Journal*, *51*(3), 42-46.

Connor, G., & Korajczyk, R. A. (1993). A test for the number of factors in an approximate factor model. *Journal of Finance*, *48*(4), 1263-1291.

Cornish, R. (2007). A comparison of the properties of realized variance for the FTSE 100 and FTSE 250 equity indices. In J. Knight & S. Satchell (Eds.). *Forecasting volatility in the financial markets* (3rd ed.). 73-100. Oxford: Elsevier Ltd.

Courtney, M. G. R. (2013). Determining the number of factors to retain in EFA: Using the SPSS R-Menu v2. 0 to make more judicious estimations. *Practical Assessment, Research and Evaluation*, *18*(8), 1-14.

Cryer, J.D., & Chan, K-S. (2008). *Time series analysis with applications in R* (2nd ed.). New York, NY: Springer.

Cutler, D.M., Poterba, J.M. & Summers, L.H. (1989). What moves stock prices? *Journal of Portfolio Management,* 15(3), 4-12.

Czaja, M. G., Scholz, H., & Wilkens, M. (2010). Interest rate risk rewards in stock returns of financial corporations: Evidence from Germany. *European Financial Management*, *16*(1), 124-154.

Davidson, R., & MacKinnon, J. G. (1981). Several tests for model specification in the presence of alternative hypotheses. *Econometrica, 49*(3), 781-793.

De Lint, C. R. (2002). Risk profiles: How do they change when stock markets collapse?. *Journal of International Financial Markets, Institutions and Money*, *12*(1), 59-80.

Deetz, M., Poddig, T., Sidorovitch, I., & Varmaz, A. (2009). An evaluation of conditional multi-factor models in active asset allocation strategies: an empirical study for the German stock market. *Financial Markets and Portfolio Management*, *23*(3), 285-313.

Desjardins, J. (2017). *The 20 largest stock exchanges in the world.* Retrieved December 18, 2017 from VisualCapitalist.com at: http://www.visualcapitalist.com/20-largest-stock-exchanges-world/

Dhrymes, P. J., Friend, I., & Gultekin, N. B. (1984). A critical reexamination of the empirical evidence on the arbitrage pricing theory. *Journal of Finance*, *39*(2), 323-346.

--- (1982). A critical reexamination of the empirical evidence on the arbitrage pricing theory. (Working Paper. No. 12-82). Wharton School, University of Pennsylvania.

Dhrymes, P. J., Friend, I., Gultekin, M. N., & Gultekin, N. B. (1985). New tests of the APT and their implications. *Journal of Finance*, *40*(3), 659-674.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*(366a), 427-431.

DiStefano, C., Zhu, M., & Mîndrilă, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation*, *14*(20), 1-11.

Dominguez, K.M. (1992). *Exchange rate efficiency and the behaviour of international asset markets.* New York, NY: Routledge

Dowd, K. (2002). *Measuring market risk* (2nd ed). Chichester: John Wiley & Sons, Inc.

Downing, D., & Clark, J. (2010). *Business Statistics* (5th ed.). Hauppauge, NY: Barron's Educational Series, Inc.

Drakos, K. (2002). Estimating a multifactor model for the Greek mutual fund market. *Russian and East European Finance and Trade*, *38*(3),73-92.

Driffill, J. (2011). The future of macroeconomics: Introductory remarks. *The Manchester School*, *79*(s2), 1-38.

Edwards, L., & Alves, P. (2006). South Africa's export performance: Determinants of export supply. *South African Journal of Economics*, *74*(3), 473-500.

Eichholtz, P. M. (1996). Does international diversification work better for real estate than for stocks and bonds? *Financial Analysts Journal*, *52*(1), 56-62.

Elkhafif, M. A.T. (1996). *The Egyptian economy: A modeling approach*. Westport, Connecticut: Praeger.

Elton, E. J., & Gruber, M. J. (1988). A multi-index risk model of the Japanese stock market. *Japan and the World Economy*, *1*(1), 21-44.

Elton, E. J., & Gruber, M. J. (1997). Modern portfolio theory, 1950 to date. *Journal of Banking and Finance*, *21*(11-12), 1743-1759.

Elton, E. J., & Gruber, M. J. (2018). The Impact of Ross's Exploration of APT on Our Research. *Journal of Portfolio Management, 44*(6), 98-107.

Elton, E. J., Gruber, M. J., & Blake, C. R. (1995). Fundamental economic variables, expected returns, and bond fund performance. *Journal of Finance*, *50*(4), 1229-1256.

Elton, E.J., Gruber, M.J., Brown, S.J. & Goetzmann, W.N. (2014). *modern portfolio theory and investment analysis* (9th ed.). New York, NY: Wiley.

Elyasiani, E. & Mansur, I. (1998). Sensitivity of the bank stock returns distribution to changes in the level and volatility of interest rate: A GARCH-M model. *Journal of Banking and Finance,* 22(5), 535-563.

Engle, R. (2001). GARCH 101: The use of ARCH/GARCH models in applied econometrics. *Journal of Economic Perspectives,* 15(4), 157-168.

Engle, R. (2004). Risk and volatility: Econometric models and financial practice. *American Economic Review*, *94*(3), 405-420.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica, 50*(4), 987-1007.

Engle, R.F., & Patton, A.J. (2007). What good is a volatility model? In J.Knight & S.Satchell (Eds.). *Forecasting volatility in the financial markets* (3rd ed.). 47 - 63. Oxford: Elsevier Ltd.

Eun, C. S., & Shim, S. (1989). International transmission of stock market movements. *Journal of Financial and Quantitative Analysis*, *24*(2), 241-256.

Fabozzi, F.J. (1998). Factor-based approach to equity portfolio management. In F.J. Fabozzi (Ed.). *Active Equity Portfolio Management*, 71-90. New Hope: Frank J. Fabozzi Associates.

Faff, R., & Chan, H. (1998). A multifactor model of gold industry stock returns: evidence from the Australian equity market. *Applied Financial Economics*, *8*(1), 21-28.

Fair, R. C. (1984). *Specification, estimation, and analysis of macroeconometric models*. Cambridge: Harvard University Press.

Fama, E. F. (1965). The behavior of stock-market prices. *Journal of Business*, *38*(1), 34-105.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, *25*(2), 383-417.

Fama, E. F. (1981). Stock returns, real activity, inflation, and money. *American Economic Review*, *71*(4), 545-565.

Fama, E. F. (1995). Random walks in stock market prices. *Financial Analysts Journal*, *51*(1), 75-80.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *Journal of Finance*, *47*(2), 427-465.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3-56.

Fama, E. F., & MacBeth, J. D. (1973). Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy*, *81*(3), 607-636.

Fan, J., Qi, L., & Xiu, D. (2014). Quasi-maximum likelihood estimation of GARCH models with heavy-tailed likelihoods. *Journal of Business and Economic Statistics*, *32*(2), 178-191.

Ferreira, M. A., & Gama, P. M. (2007). Does sovereign debt ratings news spill over to international stock markets?. *Journal of Banking and Finance*, *31*(10), 3162-3182.

Ferson, W. E., & Harvey, C. R. (1994). Sources of risk and expected returns in global equity markets. *Journal of Banking and Finance*, *18*(4), 775-803.

Fildes, R. & Kingsman, B. Incorporating demand uncertainty and forecast error in supply chain planning models. *Journal of the Operational Research Society, 62*(3), 483-500.

Fiorentini, G., Calzolari, G., & Panattoni, L. (1996). Analytic derivatives and the computation of GARCH estimates. *Journal of Applied Econometrics*, *11*(4), 399-417.

Flannery, M. J., & Protopapadakis, A. A. (2002). Macroeconomic factors do influence aggregate stock returns. *Review of Financial Studies*, *15*(3), 751-782.

Frank, M. W. (2009). Inequality and growth in the United States: Evidence from a new state-level panel of income inequality measures. *Economic Inquiry*, *47*(1), 55-68.

Frankfurter, G. M., & Lamoureux, C. G. (1990). Insignificant betas and the efficacy of the Sharpe diagonal model for portfolio selection. *Decision Sciences*, *21*(4), 853-861.

French, J. (2017). Macroeconomic Forces and Arbitrage Pricing Theory. *Journal of Comparative Asian Development, 16*(1), 1-20.

Freund, R.J., Wilson, W.J., & Sa, P. (2006). *Statistical modeling of a response variable* (2nd ed.). Burlington, MA: Elsevier Inc.

Funke, N., & Matsuda, A. (2006). Macroeconomic news and stock returns in the United States and Germany. *German Economic Review*, *7*(2), 189-210.

Gauteng Province Provincial Treasury (2013). Business cycles and their impact on the South African economy. *Quarterly Bulletin.* 1- 33. Retrieved December 4, 2017 from the Gauteng Government website at: http://www.gauteng.gov.za/government/departments/provincial-treasury/Documents/Impact%20of%20Business%20Cycles%20on%20SA%20Economy.pdf

Godfrey, L. G. (1978). Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica*, *46*(6), 1293-1301.

Granger, C. W., & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, *2*(2), 111-120.

Greene, W.J. (2012). *Econometric analysis* (7th ed.). Harlow, Essex: Pearson Education Limited.

Griffin, J. M., & Stulz, R. M. (2001). International competition and exchange rate shocks: A cross-country industry analysis of stock returns. *Review of Financial Studies*, *14*(1), 215-241.

Gujarati, D.N. (2004). *Basic Econometrics* (4th ed.). New York, NY: McGraw-Hill/Irwin.

Gujarati, D.N., & Porter, D.C. (2009). *Essentials of econometrics* (4th ed.). New York, NY: McGraw-Hill Irwin.

Gultekin, M. N., & Gultekin, N. B. (1987). Stock return anomalies and the tests of the APT. *Journal of Finance*, *42*(5), 1213-1224.

Hahn, J., & Lee, H. (2006). Yield spreads as alternative risk factors for size and book-to-market. *Journal of Financial and Quantitative Analysis*, *41*(2), 245-269.

Hamao, Y. (1988). An empirical examination of the arbitrage pricing theory: Using Japanese data. *Japan and the World Economy*, *1*(1), 45-61.

Hamilton, J.D. (2010). Macroeconomics and ARCH. In T. Bollerslev, J. Russel, & M. Watson. *Festschrift in honor of Robert F. Engle*, 79–96. Oxford: Oxford University Press.

Hanousek, J., & Kočenda, E. (2011). Foreign news and spillovers in emerging European stock markets. *Review of International Economics*, *19*(1), 170-188.

Hansen, B.E. (2012, July). *Advanced time series and forecasting - Lecture 5: structural breaks.* Lecture presented at the Summer School in Economics and Econometrics, University of Crete, Crete.

Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH (1, 1)? *Journal of Applied Econometrics*, *20*(7), 873-889.

Harris, E.S. (1991). Tracking the economy with the Purchasing Managers' Index. *FRBNY Quarterly Review (Autumn)*, 61-69.

Harvey, C. R. (1995). The risk exposure of emerging equity markets. *World Bank Economic Review*, *9*(1), 19-50.

Hassan, M. K., & Naka, A. (1996). Short-run and long-run dynamic linkages among international stock markets. *International Review of Economics and Finance*, *5*(4), 387-405.

Hauke, J., & Kossowski, T. (2011). Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae*, *30*(2), 87-93.

Theil, H. (1966). Applied economic forecasting. *American Journal of Agricultural Economics, 49*(1), 241-243.

Herwartz, H. (2004). Conditional heteroscedasticity. In H. Lütkepohl & M. Krätzig (Eds.). *Applied time series econometrics,* 197-220. Cambridge: Cambridge University Press.

Huang, W., Liu, Q., Rhee, S. G., & Zhang, L. (2009). return reversals, idiosyncratic risk, and expected returns. *Review of Financial Studies*, *23*(1), 147-168.

Hodge, V.J. & Austin, J. (2004) A survey of outlier detection methodologies. *Artificial Intelligence Review, 22*(2), 85-126.

Hughes, P. J. (1984). A test of the arbitrage pricing theory using Canadian security returns. *Canadian Journal of Administrative Sciences/Revue Canadienne des Sciences de l'Administration*, *1*(2), 195-214.

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, *127*(1), 57-83.

James, G. A., & Karoglou, M. (2010). Financial liberalization and stock market volatility: The case of Indonesia. *Applied Financial Economics*, *20*(6), 477-486.

Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, *6*(3), 255-259.

Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique*, 55(2), 163-172.

Javed, F., & Mantalos, P. (2013). GARCH-type models and performance of information criteria. *Communications in Statistics-Simulation and Computation*, *42*(8), 1917-1933.

Jennrich, R. I. (1970). An asymptotic χ2 test for the equality of two correlation matrices. *Journal of the American Statistical Association*, *65*(330), 904-912.

Jondeau, E., Poon, S-H., & Rockinger, M. (2007). *Financial modeling under non-Gaussian distributions.* London: Springer

Jorion, P. (1991). The pricing of exchange rate risk in the stock market. *Journal of Financial and Quantitative Analysis*, *26*(3), 363-376.

Kacapyr, E. (2014). *A guide to basic econometric techniques* (2nd ed.). New York, NY: Routledge

Kaiser, H.F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20(1),* 141-151.

Kauffman, R. G. (1999). Indicator Qualities of the NAPM Report On Business®. *Journal of Supply Chain Management*, *35*(1), 29-37.

Kavussanos, M. G., Marcoulis, S. N., & Arkoulis, A. G. (2002). Macroeconomic factors and international industry returns. *Applied Financial Economics*, *12*(12), 923-931.

Keim, D. B. (1986). The CAPM and equity return regularities. *Financial Analysts Journal*, *42*(3), 19-34.

Kendall, M.G. & Hill, A.B. (1953). The analysis of economic time-series-part I: Prices. *Journal of the Royal Statistical Society. Series A (General), 116*(1), 11-34.

Kim, S. J. (2003). The spillover effects of US and Japanese public information news in advanced Asia-Pacific stock markets. *Pacific-Basin Finance Journal*, *11*(5), 611-630.

King, B. F. (1966). Market and industry factors in stock price behavior. *Journal of Business*, *39*(1), 139-190.

Kluve, J., Schneider, H., Uhlendorff, A., & Zhao, Z. (2012). Evaluating continuous training programmes by using the generalized propensity score. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *175*(2), 587-617.

Konishi, S., & Kitagawa, G. (2008). *Information criteria and statistical modeling.* New Work, NY: Springer

Korajczyk, R. A., & Viallet, C. J. (1989). An empirical investigation of international asset pricing. *Review of Financial Studies*, *2*(4), 553-585.

Koutoulas, G., & Kryzanowski, L. (1994). Integration or segmentation of the Canadian stock market: Evidence based on the APT. *Canadian Journal of Economics*, *27*(2), 329-351.

Kryzanowski, L., Lalancette, S., & To, M. C. (1994). Some tests of APT mispricing using mimicking portfolios. *Financial Review*, *29*(2), 153-192.

Kryzanowski, L., & To, M. C. (1983). General factor models and the structure of security returns. *Journal of Financial and Quantitative Analysis*, *18*(1), 31-52.

Kutner, M.H., Nachtsheim, C.J., Neter, J., & Li, W. (2005). *Applied linear statistical models* (5th ed.). New York, NY: McGraw-Hill Irwin.

Kwon, O., & Yang, J. S. (2008). Information flow between composite stock index and individual stocks. *Physica A: Statistical Mechanics and its Applications*, *387*(12), 2851-2856.

Ledesma, R. D., & Valero-Mora, P. (2007). Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. *Practical Assessment, Research & Evaluation*, *12*(2), 1-11.

Lee, C. G. (2011). The determinants of hotel room rates: Another visit with Singapore's data. *International Journal of Hospitality Management*, *30*(3), 756-758.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48*(2), 281-355.

Lehmann, B. N. (1990). Residual risk revisited. *Journal of Econometrics*, *45*(1-2), 71-97.

Lehmann, B. N., & Modest, D. M. (1987). Mutual fund performance evaluation: A comparison of benchmarks and benchmark comparisons. *Journal of Finance*, *42*(2), 233-265.

Leightner, J. E., & Inoue, T. (2012). Solving the omitted variables problem of regression analysis using the relative vertical position of observations. *Advances in Decision Sciences*, *1-26*

Lim, J. (2003). Political risk and the exchange rate: An exploration with a regime switching model. *Working Paper.* Retrieved December 4, 2017 from the Academia.edu website: http://www.academia.edu/download/30304340/ipe3.pdf.

Lin, C. Y., Rahman, H., & Yung, K. (2009). Investor sentiment and REIT returns. *Journal of Real Estate Finance and Economics*, *39*(4), 450-471.

Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. *Journal of Finance*, *20*(4), 587-615.

Liow, K., & Ibrahim, M. F., & Huang, Q. (2006). Macroeconomic risk influences on the property stock market. *Journal of Property Investment and Finance*, *24*(4), 295-323.

Ljung, G.M., & Box, G.P.E. (1978). On a measure of lack of fit in time series models. *Biometrika, 65*(2), 297-303.

Lo, A.W. (2004). The adaptive market hypothesis. *Journal of Portfolio Management, 30*(5), 15-29.

Renault, E., van der Heijden, T., & Werker, B. (2016). *Arbitrage pricing theory for idiosyncratic variance factors.* SSRN Working Paper. Retrieved November 6, 2018 from the SSRN website: http://dx.doi.org/10.2139/ssrn.3065854

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, *17*(1), 59-82.

Masih, A. M., & Masih, R. (1999). Are Asian stock market fluctuations due mainly to intra-regional contagion effects? Evidence based on Asian emerging stock markets. *Pacific-Basin Finance Journal*, *7*(3-4), 251-282.

Mateus, T. (2004). The risk and predictability of equity returns of the EU accession countries. *Emerging Markets Review*, *5*(2), 241-266.

McDonald, R.P. (1985). *Factor analysis and related methods*. Hillside, NJ: Lawrence Erlbaum Associates, Inc.

McElroy, M. B., & Burmeister, E. (1988). Arbitrage pricing theory as a restricted nonlinear multivariate regression model iterated nonlinear seemingly unrelated regression estimates. *Journal of Business and Economic Statistics*, *6*(1), 29-42.

McQueen, G., & Roley, V. V. (1993). Stock prices, news, and business conditions. *Review of Financial Studies*, *6*(3), 683-707.

Mela, C. F., & Kopalle, P. K. (2002). The impact of collinearity on regression analysis: the asymmetric effect of negative and positive correlations. *Applied Economics*, *34*(6), 667-677.

Meyers, S. L. (1973). A re-examination of market and industry factors in stock price behavior. *Journal of Finance*, *28*(3), 695-705.

Middleton, L. P., & Satchell, S. E. (2001). Deriving the arbitrage pricing theory when the number of factors is unknown. *Quantitative Finance*, *1*(5), 502-508.

Mills, T.C., & Markellos, R.N. (2008). *The Econometric Modelling of Financial Time Series* (3rd ed.). Cambridge: Cambridge University Press.

Mittelhammer, R. C., Judge, G.G., & Miller, D.J. (2000). *Econometric foundations*. Cambridge: Cambridge University Press.

Moolman, E. (2003). Predicting turning points in the South African economy. *South African Journal of Economic and Management Sciences, 6*(2), 289-303.

MSCI (2017). *MSCI Index Calculation Methodology.* Retrieved December 1, 2017 from the MSCI website at: https://www.msci.com/eqb/methodology/meth_docs/MSCI_July2017_IndexCalcMethodology.pdf

MSCI (2018). MSCI World Index (USD) Fund Factsheet. Retrieved June 1, 2018 from the MSCI website at: https://www.msci.com/documents/10199/178e6643-6ae6-47b9-82be-e1fc565ededb

Muradoglu, G., Taskin, F., & Bigan, I. (2000). Causality between stock returns and macroeconomic variables in emerging markets. *Russian and East European Finance and Trade, 36*(6), 33-53.

Mutsune, T. (2008). The state of US international competitiveness: A study of the impact of trade performance indicators. *Advances in Competitiveness Research, 16*(1/2), 1-12.

Naifar, N., & Al Dohaiman, M. S. (2013). Nonlinear analysis among crude oil prices, stock markets' return and macroeconomic variables. *International Review of Economics & Finance*, *27*, 416-431.

Nel, W. S. (2011). The application of the capital asset pricing model (CAPM): A South African perspective. *African Journal of Business Management*, *5*(13), 5336.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica, 59*(2), 347-370.

Nelson, D., & Cao, C. (1992). Inequality Constraints in the Univariate GARCH Model. *Journal of Business & Economic Statistics, 10*(2), 229-235

Newey, W. K., & West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, *28*(3), 777-787.

Niemira, M.P. (1993). Using composite leading indicators of consumption to forecast sales and to signal turning points in the stock market. In K. Lahiri, & G.H. Moore (Eds.). *Leading economic indicators: New approaches and forecasting records.* Cambridge: Cambridge University Press.

Nwogugu, M. (2006). Further critique of GARCH/ARMA/VAR/EVT stochastic-volatility models and related approaches. *Applied Mathematics and Computation*, *182*(2), 1735–1748.

Onwuegbuzie, A. J., & Daniel, L. G. (2002). Uses and misuses of the correlation coefficient. *Research in the Schools, 9*(1), 73-90.

Osamwonyi, I. O., & Evbayiro-Osagie, E. I. (2012). The relationship between macroeconomic variables and stock market index in Nigeria. *Journal of Economics*, *3*(1), 55-63.

Pan, M. S., Fok, R. C. W., & Liu, Y. A. (2007). Dynamic linkages between exchange rates and stock prices: Evidence from East Asian markets. *International Review of Economics & Finance*, *16*(4), 503-520.

Panetta, F. (2002). The stability of the relation between the stock market and macroeconomic forces. *Economic Notes*, *31*(3), 417-450.

Partalidou, X., Kiohos, A., Giannarakis, G., & Sariannidis, N. (2016). The impact of gold, bond, currency, metals and oil markets on the USA stock market. *International Journal of Energy Economics and Policy*, *6*(1), 76-81.

Peavler, R. (2018). *2001 to 2009: Top 10 financial events of the decade.* Retrieved May 30, 2018 from the balance small business website at: https://www.thebalancesmb.com/top-10-financial-events-of-the-decade-393162

Phillips, P. C., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, *75*(2), 335-346.

Pilinkus, D. (2010). Macroeconomic indicators and their impact on stock market performance in the short and long run: the case of the Baltic States. *Baltic Journal of Sustainability*, *16*(2), 291-304.

Poon, S-H. (2005). *A practical guide to forecasting financial market volatility.* Chichester: John Wiley & Sons, Inc.

Poon, S., & Taylor, S. J. (1991). Macroeconomic factors and the UK stock market. *Journal of Business Finance and Accounting*, *18*(5), 619-636.

Priestley, R. (1996). The arbitrage pricing theory, macroeconomic and financial factors, and expectations generating processes. *Journal of Banking and Finance*, *20*(5), 869-890.

PWC (2015). Africa: A closer look at value. Retrieved December 1, 2017, from the PWC website at https://www.pwc.co.za/en/assets/pdf/valuation-methodology-survey-2015.pdf

Quantitative Mico Software (2009). *Eviews 7 user's guide II.* Irvine, CA: QMS, LLC

Ramsey, J. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society. Series B (Methodological), 31*(2), 350-371.

Reilley, F.K., & Brown, C. (2012). *Investment analysis & portfolio management* (10th ed.). Mason, USA: South Western Cengage Learning.

Reinganum, M.R. (1981). The arbitrage pricing theory: Some empirical results. *Journal of Finance, 36*(2), 313-321.

Roberts, M. R., & Whited, T. M. (2013). Endogeneity in empirical corporate finance. In G.M. Constantinides, M. Harris & R.M. Stulz (Eds.). *Handbook of the Economics of Finance,* 2B, 493-572. Oxford: Elsevier.

Roll, R. (1992). Industrial structure and the comparative behavior of international stock market indices. *Journal of Finance, 47*(1), 3-41.

Roll, R., & Ross, S. A. (1980). An empirical investigation of the arbitrage pricing theory. *Journal of Finance*, *35*(5), 1073-1103.

Roll, R., & Ross, S. A. (1995). The arbitrage pricing theory approach to strategic portfolio planning. *Financial Analysts Journal, 51(*1), 122-131.

Ross, S.A., (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3), 341-360.

Rummel, R. J. (1967). Understanding factor analysis. *Journal of Conflict Resolution*, *11*(4), 444-480.

Sadorsky, P. (2001). Risk factors in stock returns of Canadian oil and gas companies. *Energy Economics*, *23*(1), 17-28.

Sadorsky, P. (2008). Assessing the impact of oil prices on firms of different sizes: Its tough being in the middle. *Energy Policy*, *36*(10), 3854-3861.

Sadorsky, P., & Henriques, I. (2001). Multifactor risk and the stock returns of Canadian paper and forest products companies. *Forest Policy and Economics*, *3*(3-4), 199-208.

Serre, D. (2002). *Matrices: Theory and applications.* New York, NY: Springer.

Sharpe, W. F. (1963). A simplified model for portfolio analysis. *Management Science*, *9*(2), 277-293.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, *19*(3), 425-442.

Singh, P., Kumar, B., & Pandey, A. (2010). Price and volatility spillovers across North American, European and Asian stock markets. *International Review of Financial Analysis*, *19*(1), 55-64.

Smith, V. K., & Hall, T. W. (1972). A comparison of maximum likelihood versus BLUE estimators. *Review of Economics and Statistics*, *54*(2), 186-190.

Solnik, B. (1983). International arbitrage pricing theory. *Journal of Finance*, *38*(2), 449-457.

Solnik, B. H. (1974). The international pricing of risk: An empirical investigation of the world capital market structure. *Journal of Finance*, *29*(2), 365-378.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(3), 485-493.

Spronk, J., & Hallerbach, W. (1997). Financial modelling: Where to go? With an illustration for portfolio management. *European Journal of Operational Research*, *99*(1), 113-125.

Spyridis, T., Sevic, Z., & Theriou, N. (2012). Macroeconomic vs. statistical APT approach in the Athens Stock Exchange. *International Journal of Business*, *17*(1), 39-64.

Statistics South Africa (2017). Retail trade sales (Preliminary). Retrieved May 1, 2017 from the Statistics South Arica website at: http://www.statssa.gov.za/publications/P62421/P62421December2016.pdf

Studenmund, A.H. (2014). *Using econometrics: a practical guide* (6th ed.). Harlow, Essex: Pearson Education Limited.

Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the P value is not enough. *Journal of Graduate Medical Education*, *4*(3), 279-282.

Sweeney, R. J., & Warga, A. D. (1986). The pricing of interest-rate risk: evidence from the stock market. *Journal of Finance*, *41*(2), 393-410.

Sykes, A. O. (1993). *An introduction to regression analysis (Working Paper No. 20)*. Coase-Sandor Institute for Law and Economics, University of Chicago Law School.

Szczygielski, J. J., & Chipeta, C. (2015). Risk factors in returns of the South African stock market. *Studies in Economics and Econometrics*, *39*(1), 47-70.

Taylor, S.J. (2008). *Modelling financial time series* (2nd ed). Singapore: World Scientific Publishing Co. Pty. Ltd.

Thadewald, T., & Büning, H. (2007). Jarque–Bera test and its competitors for testing normality–a power comparison. *Journal of Applied Statistics*, *34*(1), 87-105.

Thorbecke, W. (1997). On stock market returns and monetary policy. *Journal of Finance, 52(2), 635-654.*

Treynor, J. L. (1961). Toward a theory of market value of risky assets. *Unpublished manuscript*, *6*, 831-868.

Tsay, R.S. (2002). *Analysis of financial time series.* New York: John Wiley & Sons, Inc.

Van Rensburg, P. (1995). Macroeconomic variables and the Johannesburg Stock Exchange: A multifactor approach. *De Ratione*, *9*(2), 45-63.

Van Rensburg, P. (1996). Macroeconomic identification of the priced APT factors on the Johannesburg Stock Exchange. *South African Journal of Business Management*, *27*(4), 104-112.

Van Rensburg, P. (1997). Employing the prespecified variable approach to APT factor identification on the segmented Johannesburg Stock Exchange. *South African Journal of Accounting Research*, *11*(1), 57-74.

Van Rensburg, P. (1998). Unifying the factor analytic and prespecified variable approaches to APT factor identification on the Johannesburg Stock Exchange. *South African Journal of Accounting Research, 12*(1), 15-45.

Van Rensburg, P. (1999). Macroeconomic identification of candidate APT factors on the Johannesburg Stock Exchange. *Journal for Studies in Economic and Econometrics*, 23, 27-53.

Van Rensburg, P. (2000). Macroeconomic variables and the cross-section of Johannesburg Stock Exchange returns. *South African Journal of Business Management, 31*(1), 31-43.

Van Rensburg, P. (2002). Market segmentation on the Johannesburg stock exchange II. *Studies in Economics and Econometrics*, *26*(1), 83-99.

Van Rensburg, P., & Slaney K. (1997). Market segmentation on the Johannesburg Stock Exchange. *Journal for Studies in Economics and Econometrics,* 23(3), 1-23.

Varga, J., & Rappai, G. (2002). Heteroscedasticity and efficient estimates of beta. *Hungarian Statistical Review*, *80*(7), 127-137.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321-327.

Venter, J.C. (2005). Reference turning points in the South African business cycle: Recent developments. *SARB Quarterly Bulletin,* 61-70.

Venter, J. C., & Pretorius, W. S. (2004). Note on the revision of composite leading and coincident business cycle indicators. *SARB Quarterly Bulletin*, 67-72.

Vogt, M. G. (2008). Determinants of the demand for US exports and imports of tourism. *Applied Economics*, *40*(6), 667-672.

Walker, J & Madden, S. (2008). *Statistics in criminology and criminal justice: Analysis and interpretation* (3rd ed.). Sadbury, Massachusetts: John Bartlett Publishers.

Walther, B.W, and Moore, J.L. (2005). The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography, 28*, 815-829.

Watson, P.K., & Teelucksingh, S. S. (2002). *A practical introduction to econometric methods: classical and modern.* Kingston: University of the West Indies Press

Webster, A. (2013). *Introductory regression analysis: With computer application for business and economics*. New York, NY: Routledge.

Wei, K. J. (1988). An asset-pricing theory unifying the capm and apt. *Journal of Finance*, *43*(4), 881-892.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica, 48*(4), 817-838.

Williams, M. N., Grajales, C. A. G., & Kurkiewicz, D. (2013). assumptions of multiple regression: Correcting two misconceptions. *Practical Assessment, Research & Evaluation*, *18*(11),1-14.

Wong, W-K., & Bian, G. (2000). Robust estimation in capital asset pricing model. *Journal of Applied Mathematics and Decision Sciences, 4*(1)*, 65-82.*

Wongbangpo, P., & Sharma, S. C. (2002). Stock market and macroeconomic fundamental dynamic interactions: ASEAN-5 countries. *Journal of Asian Economics*, *13*(1), 27-51.

Wooldridge, J.M. (2013). *Introductory econometrics: a modern approach* (5th ed.). Mason, OH: South-Western CENGAGE Learning

Wurm, L. H., & Fisicaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language*, *72*, 37-48.

Xiao, L. & Aydemir, A. (2007). Volatility modeling and forecasting in finance. In J. Knight & S. Satchell (Eds.). *Forecasting volatility in the financial markets* (3rd ed.). 1-45. Oxford: Elsevier Ltd.

Yao, W., Mei, B., & Clutter, M. L. (2014). Pricing timberland assets in the United States by the arbitrage pricing theory. *Forest Science*, *60*(5), 943-952.

Yli-Olli, P., & Virtanen, I. (1992). Some empirical tests of the arbitrage pricing theory using transformation analysis. *Empirical Economics*, *17*(4), 507-522.

Yohai, V. (1987). High breakdown-point and high efficiency robust estimates for regression. *Annals of Statistics, 15*(2), 642-656.

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology*, *9*(2), 79-94.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*(3), 432-442.

APPENDIX A: UNABRIDGED RESULTS

Table A1.1., Table A.1.2., Table A1.3. and Table A1.4. report the unabridged results of the benchmark model, the restricted model, the unrestricted market model and the unrestricted model respectively.

In Panel A, coefficients are reported, together with the standard errors (in parentheses) and the associated *z*-scores [in brackets]. The asterisks indicate statistical significance, at the 1% level of significance (***) at the 5% level of significance (**) and at the 10% level of significance (*). Panel B reports the goodness-of–fit measures and model selection criteria. These are the adjusted coefficient of determination, $\bar{R}^2$, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

Panel C reports the model diagnostics. These are the *F*-statistics for Wald's test of linear restrictions restricting all coefficients to zero in a given specification, test statistics for the Jarque-Bera test of normality (JB), Ljung-Box Q-statistics for residual serial correlation at the first order, $Q(1)$ and joint serial correlation up to the fifth order of serial correlation, $Q(5)$. Squared *Q*-statistics are reported for tests for non-linear dependence in the residuals at the first order, $Q^2(1)$, and jointly up to the fifth order, $Q^2(5)$. ARCH(1) and ARCH(5) are test statistics for the ARCH LM test for ARCH effects at the first and fifth orders. The asterisks indicate statistical significance at the 1% level of significance (***), at the 5% level of significance (**) and at the 10% level of significance (*).

The residual variance, $\sigma^2_{\varepsilon i}$, for each sector derived for each specification is reported in Panel D. In Panel D of Table A1.2., Table A1.3., and Table A1.4., the first set of ▲ or ▼ symbols indicates whether residual variance is greater or lower relative to that for the benchmark model. The second set in Table A1.3. and Table A1.4. indicates whether the residual variance is greater or lower relative to that for the restricted model. A single asterisk (*) is used to denote statistically significant differences at the 10% level of significance (to conserve space) for the Brown-Forsythe test of the equality of variances.

Panel E reports the results of the ARCH*(p)* and GARCH (*p,q*) models used to describe the conditional variance structure underlying each industrial sector modelled using each specification of the linear factor model. *F*-statistics are for Wald's test of linear restrictions, restricting all ARCH and GARCH coefficients to zero in the respective ARCH(*p*) or GARCH(*p,q*) specifications.

Panel F reports the means of the residuals for each sector derived from each specification of the linear factor model. Theil U statistics are reported in Panel G, together with the associated components, namely the bias proportion, U_{BIAS}, the variance proportion, U_{VAR}, and the covariance proportion, U_{COV} .

Finally, the results of the likelihood ratio tests for omitted factors are reported in Panel H of Table A1.2., Table A1.3. and Table A1.4. Omitted factors are hypothesised to be $M\varepsilon_t$, $IM\varepsilon_t$ and f_{1t},f_{2t}. The asterisks indicate statistical significance at the 1% level of significance (***), at the 5% level of significance (**) and at the 10% level of significance (*).

.

Table A1.1: Benchmark Model Results

	Chemicals	Forestry & Paper	Ind. Met & Mining	Mining	Const & Materials	General Indust.	Elect. & Elect. Eq.	Industrial Eng.	Industrial Transport.
Panel A: Model Results									
Intercept	0.007	0.003	0.008	0.000	0.004	0.009	0.001	0.010	0.004
	(0.003)	(0.005)	(0.005)	(0.001)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)
	[2.306]**	[0.49]	[1.459]	[0.217]	[1.182]	[4.335]***	[0.339]	[3.011]***	[1.54]
BP_{t-1}	0.047	0.023	0.058	0.046	0.035	0.056	0.025	-0.001	0.069
	(0.028)	(0.046)	(0.056)	(0.014)	(0.04)	(0.022)	(0.026)	(0.027)	(0.023)
	[1.677]**	[0.505]	[1.027]	[3.349]***	[0.868]	[2.506]**	[0.962]	[-0.032]	[2.96]***
$LEAD_{t-1}$	0.842	0.740	0.435	1.178	0.382	1.173	0.461	1.750	1.688
	(0.363)	(0.654)	(0.627)	(0.185)	(0.431)	(0.295)	(0.37)	(0.435)	(0.287)
	[2.317]**	[1.131]	[0.694]	[6.367]***	[0.887]	[3.978]***	[1.247]	[4.025]***	[5.875]***
BUS_t	0.019	0.131	0.055	0.003	0.106	0.072	0.097	0.115	0.112
	(0.031)	(0.058)	(0.052)	(0.016)	(0.047)	(0.022)	(0.033)	(0.041)	(0.035)
	[0.635]	[2.24]**	[1.072]	[0.189]	[2.243]**	[3.297]***	[2.943]***	[2.796]***	[3.174]***
$USD\varepsilon_t$	-0.277	0.246	-0.550	0.153	-0.515	-0.163	-0.311	-0.168	-0.159
	(0.095)	(0.168)	(0.193)	(0.052)	(0.119)	(0.086)	(0.093)	(0.107)	(0.088)
	[-2.917]***	[1.464]	[-2.845]***	[2.923]***	[-4.324]***	[-1.898]*	[-3.326]***	[-1.566]	[-1.813]*
MET_t	0.217	0.353	0.697	0.461	0.172	0.01	0.34	0.41	0.046
	(0.055)	(0.108)	(0.11)	(0.043)	(0.07)	(0.055)	(0.065)	(0.084)	(0.06)
	[3.934]***	[3.275]***	[6.342]***	[10.632]***	[2.444]**	[0.174]	[5.198]***	[4.906]***	[0.764]
LTY_t	-3.029	-2.573	-0.189	-0.116	-4.969	-4.689	-3.944	-4.833	-7.386
	(0.791)	(1.603)	(1.666)	(0.498)	(1.261)	(0.862)	(0.905)	(1.152)	(0.859)
	[-3.829]***	[-1.605]	[-0.113]	[-0.232]	[-3.94]***	[-5.438]***	[-4.357]***	[-4.195]***	[-8.593]***
$TLI\varepsilon_t$	2.567	6.124	3.673	5.192	1.513	2.430	2.035	2.546	2.924
	(0.657)	(1.262)	(1.32)	(0.46)	(0.883)	(0.584)	(0.697)	(0.794)	(0.618)
	[3.907]***	[4.854]***	[2.782]***	[11.289]***	[1.713]*	[4.165]***	[2.921]***	[3.206]***	[4.733]***
$M\varepsilon_t$	0.464	1.053	1.192	1.409	0.591	0.696	0.535	0.546	0.668
	(0.079)	(0.193)	(0.124)	(0.035)	(0.083)	(0.055)	(0.065)	(0.117)	(0.062)
	[5.86]***	[5.458]***	[9.615]***	[40.101]***	[7.15]***	[12.752]***	[8.243]***	[4.659]***	[10.726]***
$IM\varepsilon_t$	0.300	0.457	0.242	-0.457	0.264	0.047	0.328	0.416	0.243
	(0.081)	(0.201)	(0.176)	(0.05)	(0.138)	(0.079)	(0.106)	(0.129)	(0.103)
	[3.685]***	[2.271]**	[1.369]	[-9.121]***	[1.917]**	[0.596]	[3.1]***	[3.225]***	[2.354]**
f_1	-	-	-0.055***	-0.0359***	-	0.015***	-	-0.009***	-
f_2	0.021***	-	0.024***	-	0.032***	0.013***	0.020***	0.019***	0.030***

Table A1.1: Benchmark Model Results (Continued…)

	Chemicals	Forestry & Paper	Ind. Met & Mining	Mining	Const & Materials	General Indust.	Elect. & Elect. Eq.	Industrial Eng.	Industrial Transport.
Panel B: Goodness-of-fit And Model Selection Criteria									
$\bar{R}^2$	0.451	0.373	0.602	0.941	0.529	0.621	0.575	0.469	0.692
AIC	-3.550	-2.343	-2.224	-4.956	-3.182	-4.038	-3.753	-3.093	-3.747
BIC	-3.313	-2.140	-1.969	-4.736	-2.962	-.3.800	-3.532	-2.855	-3.527
Panel C: Model Diagnostics									
F-Statistic	26.463***	12.538***	17.677***	228.541***	30.030***	24.917***	32.870***	10.576***	75.526***
JB	8.268**	111.207***	10.044***	4.774*	13.516***	4.772*	2.508	39.439***	4.085
$Q(1)$	3.882**	0.246	0.832	0.273	1.033	1.734	0.815	2.948*	1.917
$Q(5)$	8.114	7.798	4.129	5.551	8.059	5.981	5.556	15.445***	7.820
$Q^2(1)$	0.523	0.029	1.378	0.010	0.041	0.327	0.041	0.048	0.064
$Q^2(5)$	7.090	2.143	4.046	1.734	2.355	6.516	1.102	2.055	2.561
ARCH(1)	0.508	0.028	1.348	0.010	0.040	0.319	0.040	0.046	0.062
ARCH(5)	1.511	0.384	0.719	0.318	0.788	1.497	0.235	0.363	0.499
Panel D: Residual Variance									
$\sigma^2_{\varepsilon i}$	0.001375	0.006090	0.001974	0.000804	0.001952	0.003567	0.001641	0.000907	0.003399
Panel E: Conditional Variance Structure (ARCH/GARCH)									
ω	2.11E-05	0.005***	1.86E-03***	8.38E-04***	0.000	0.003***	0.002***	0.001***	0.0001
α_1	0.066	0.236**	0.054	-0.049	0.059	0.178	-0.028	0.091	0.075
β_1	0.930***				0.794***	-			0.850***
F-Statistic	1314.531***	3.924**	0.786	0.367	56.722***	2.579	0.144	0.599	157.808***
Panel F: Mean Errors									
ε_{it}	-0.0006127	0.000174	-0.0000323	0.000552	-0.0020306	-0.0002546	-0.0000384	-0.0004609	.0000745
Panel G: Theil's *U* And Decomposition									
Theil *U*	0.419	0.473	0.352	0.120	0.384	0.321	0.358	0.407	0.293
U_{BIAS}	0.000239	0.000016	0.000000	0.000832	0.001909	0.000071	0.000001	0.000091	0.000004
U_{VAR}	0.164528	0.225970	0.190431	0.022184	0.152868	0.078910	0.126385	0.167293	0.084426
U_{COV}	0.835234	0.774013	0.809569	0.976984	0.845224	0.921019	0.873614	0.832616	0.915570

Table A1.1: Benchmark Model Results (Continued…)

	Support Services	Automobiles & Parts	Beverages	Food Producers	Health Care Eq. & Serv.	Pharm. & Biotech.	Food & Drug Retail.	General Retailers	Media
Panel A: Model Results									
Intercept	0.002	0.006	0.009	0.009	0.011	0.013	0.009	0.009	0.014
	(0.002)	(0.006)	(0.003)	(0.002)	(0.003)	(0.004)	(0.003)	(0.002)	(0.004)
	[0.628]	[1.105]	[2.781]***	[3.976]***	[3.834]***	[3.215]***	[2.891]***	[3.758]***	[3.726]***
BP_{t-1}	0.054	-0.018	0.010	0.025	0.016	-0.008	0.117	0.088	0.053
	(0.022)	(0.06)	(0.036)	(0.018)	(0.025)	(0.036)	(0.029)	(0.026)	(0.033)
	[2.443]**	[-0.301]	[0.274]	[1.383]	[0.638]	[-0.235]	[4.033]***	[3.325]***	[1.609]
$LEAD_{t-1}$	0.727	1.237	1.124	0.633	0.721	1.337	0.553	1.124	0.877
	(0.299)	(0.614)	(0.445)	(0.283)	(0.379)	(0.573)	(0.454)	(0.289)	(0.438)
	[2.431]**	[2.015]**	[2.527]**	[2.24]**	[1.903]*	[2.333]**	[1.218]	[3.887]***	[2.002]**
BUS_t	0.092	0.080	0.047	0.080	0.064	0.110	0.156	0.081	0.143
	(0.032)	(0.071)	(0.039)	(0.025)	(0.038)	(0.049)	(0.038)	(0.026)	(0.042)
	[2.887]***	[1.125]	[1.217]	[3.172]***	[1.669]*	[2.225]**	[4.054]***	[3.093]***	[3.444]***
$USD\varepsilon_t$	-0.239	-0.024	0.060	-0.072	-0.176	-0.434	-0.297	-0.429	-0.111
	(0.077)	(0.184)	(0.102)	(0.07)	(0.113)	(0.162)	(0.108)	(0.082)	(0.127)
	[-3.121]***	[-0.13]	[0.593]	[-1.026]	[-1.566]	[-2.682]***	[-2.753]***	[-5.214]***	[-0.873]
MET_t	0.065	0.164	-0.068	0.083	0.015	0.036	-0.182	0.008	0.005
	(0.056)	(0.112)	(0.073)	(0.045)	(0.084)	(0.083)	(0.065)	(0.047)	(0.08)
	[1.174]	[1.47]	[-0.937]	[1.862]*	[0.176]	[0.431]	[-2.812]***	[0.176]	[0.058]
LTY_t	-4.477	-3.469	1.530	-4.812	-5.173	-6.148	-4.286	-8.665	-4.631
	(0.667)	(2.064)	(1.221)	(0.667)	(0.975)	(1.647)	(1.128)	(0.714)	(1.104)
	[-6.717]***	[-1.681]*	[1.253]	[-7.211]***	[-5.306]***	[-3.734]***	[-3.799]***	[-12.129]***	[-4.194]***
$TLI\varepsilon_t$	2.369	4.505	3.089	2.556	2.875	0.904	0.423	1.655	5.520
	(0.566)	(1.398)	(0.796)	(0.555)	(0.693)	(1.165)	(0.71)	(0.504)	(1.011)
	[4.183]***	[3.223]***	[3.879]***	[4.604]***	[4.147]***	[0.776]	[0.596]	[3.285]***	[5.461]***
$M\varepsilon_t$	0.500	0.413	0.795	0.449	0.499	0.424	0.312	0.569	0.898
	(0.056)	(0.121)	(0.083)	(0.052)	(0.072)	(0.097)	(0.084)	(0.065)	(0.094)
	[8.885]***	[3.4]***	[9.625]***	[8.674]***	[6.951]***	[4.373]***	[3.699]***	[8.737]***	[9.502]***
$IM\varepsilon_t$	0.246	0.261	0.009	-0.009	-0.057	0.328	0.098	0.279	0.305
	(0.077)	(0.212)	(0.115)	(0.083)	(0.106)	(0.169)	(0.126)	(0.07)	(0.146)
	[3.205]***	[1.232]	[0.082]	[-0.114]	[-0.539]	[1.943]**	[0.773]	[4.002]***	[2.086]**
f_1	0.013***	-	-0.014***	0.008***	0.013***	0.013***	0.016***	0.025***	0.027***
f_2	0.020***	0.028***	-	0.015***	0.018***	0.023***	0.022***	0.029***	0.009***

Table A1.1: Benchmark Model Results (Continued…)

	Support Services	Automobiles & Parts	Beverages	Food Producers	Health Care Eq. & Serv.	Pharm. & Biotech.	Food & Drug Retail.	General Retailers	Media
Panel B: Goodness-of-fit And Model Selection Criteria									
$\bar{R}^2$	0.533	0.205	0.428	0.565	0.402	0.284	0.481	0.779	0.519
AIC	-3.795	-2.155	-3.260	-4.153	-3.334	-2.684	-3.436	-4.033	-2.826
BIC	-3.541	-1.935	-3.040	-3.916	-3.080	-2.446	-3.198	-3.795	-2.572
Panel C: Model Diagnostics									
F-Statistic	37.410***	9.020***	13.539***	26.271***	13.036***	11.052***	13.335***	123.028***	23.370***
JB	1.388	22.337***	11.182***	3.447	5.012*	11.837***	5.022*	1.362	6.037**
$Q(1)$	0.373	0.022	2.222	2.645	2.136	1.158	0.730	0.0003	0.005
$Q(5)$	3.492	3.339	8.748	4.373	5.304	7.871	4.103	3.700	5.006
$Q^2(1)$	0.640	0.245	0.067	0.188	0.079	0.126	0.675	0.002	0.070
$Q^2(5)$	2.246	3.826	2.463	1.8550	7.856	3.335	3.446	5.054	2.230
ARCH(1)	0.624	0.238	0.065	0.188	0.077	0.122	0.775	0.002	0.068
ARCH(5)	0.424	0.603	0.482	0.315	1.603	0.565	0.817	1.139	0.421
Panel D: Residual Variance									
$\sigma^2_{\varepsilon i}$	0.001375	0.006090	0.001974	0.000804	0.001952	0.003567	0.001641	0.000907	0.003399
Panel E: Conditional Variance Structure (ARCH/GARCH)									
ω	2.11E-05	0.005***	1.86E-03***	8.38E-04***	0.000	0.003***	0.002***	0.001***	0.0001
α_1	0.066	0.236**	0.054	-0.049	0.059	0.178	-0.028	0.091	0.075
β_1	0.930***				0.794***	-			0.850***
F-statistic	1314.531***	3.924**	0.786	0.367	56.722***	2.579	0.144	0.599	157.808***
Panel F: Mean Errors									
ε_{it}	-0.001236	-0.0028085	-0.0006243	0.0001953	0.0019807	0.0005116	0.0005493	-0.0000978	0.0001327
Panel G: Theil's *U* And Decomposition									
Theil *U*	0.382	0.572	0.433	0.353	0.440539	0.498	0.401	0.239	0.392
U_{BIAS}	0.001115	0.001300	0.000198	0.000048	0.002017	0.000074	0.000185	0.000011	0.000005
U_{VAR}	0.161703	0.316929	0.197865	0.130016	0.209046	0.2297944	0.167177	0.060884	0.219606
U_{COV}	0.837182	0.681771	0.460928	0.869936	0.788937	0.771982	0.832638	0.939105	0.780388

Table A1.1: Benchmark Model Results (Continued…)

	Travel & Leisure	Fixed Line Telecom.	Banks	Non-life Insurance	Life Insurance	General Financial	Equity Inv. & Inst.	Soft. & Com Serv.
Panel A: Model Results								
Intercept	0.006	0.003	0.004	0.005	0.002	0.003	0.005	0.006
	(0.003)	(0.006)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.004)
	[2.032]**	[0.456]	[1.879]*	[1.629]	[0.673]	[1.327]	[2.252]***	[1.433]
BP_{t-1}	-0.006	0.058	0.047	-0.009	0.019	0.051	0.062	0.049
	(0.027)	(0.064)	(0.023)	(0.031)	(0.023)	(0.023)	(0.024)	(0.043)
	[-0.219]	[0.9]	[2.024]**	[-0.283]	[0.813]	[2.17]***	[2.592]***	[1.133]
$LEAD_{t-1}$	0.230	0.786	0.656	0.359	1.432	1.485	0.301	1.308
	(0.353)	(0.84)	(0.324)	(0.46)	(0.288)	(0.348)	(0.274)	(0.575)
	[0.653]	[0.935]	[2.022]**	[0.78]	[4.972]***	[4.274]***	[1.096]	[2.273]**
BUS_t	0.042	0.018	0.107	0.078	0.059	0.091	0.054	0.053
	(0.031)	(0.059)	(0.032)	(0.034)	(0.028)	(0.032)	(0.028)	(0.052)
	[1.34]	[0.309]	[3.31]***	[2.283]**	[2.088]**	[2.873]***	[1.912]*	[1.017]
$USD\varepsilon_t$	-0.142	-0.061	-0.283	-0.230	-0.178	-0.223	0.048	-0.144
	(0.087)	(0.201)	(0.093)	(0.113)	(0.082)	(0.095)	(0.104)	(0.152)
	[-1.645]*	[-0.301]	[-3.025]***	[-2.043]**	[-2.168]**	[-2.334]***	[0.465]	[-0.951]
MET_t	0.097	0.279	0.068	0.137	0.151	0.214	0.01	0.231
	(0.068)	(0.128)	(0.058)	(0.081)	(0.042)	(0.056)	(0.057)	(0.111)
	[1.421]	[2.181]**	[1.177]	[1.691]*	[3.591]***	[3.8]***	[0.18]	[2.081]**
LTY_t	-4.572	-2.240	-4.339	-4.364	-5.243	-4.638	-0.925	-3.733
	(0.789)	(2.775)	(0.889)	(1.067)	(0.803)	(0.983)	(0.965)	(1.467)
	[-5.795]***	[-0.807]	[-4.88]***	[-4.09]***	[-6.528]***	[-4.717]***	[-0.959]	[-2.544]**
$TLI\varepsilon_t$	2.530	2.814	1.713	2.136	4.599	2.345	1.854	3.586
	(0.556)	(1.593)	(0.665)	(1.006)	(0.596)	(0.587)	(0.506)	(1.295)
	[4.555]***	[1.766]*	[2.575]***	[2.123]**	[7.711]***	[3.998]***	[3.663]***	[2.769]***
$M\varepsilon_t$	0.568	0.656	0.724	0.459	0.724	0.767	0.592	0.769
	(0.075)	(0.141)	(0.062)	(0.093)	(0.065)	(0.063)	(0.062)	(0.114)
	[7.568]***	[4.638]***	[11.765]***	[4.936]***	[11.135]***	[12.14]***	[9.51]***	[6.733]***
$IM\varepsilon_t$	0.336	0.403	0.284	0.058	0.600	0.399	0.122	0.137
	(0.103)	(0.244)	(0.092)	(0.12)	(0.084)	(0.102)	(0.092)	(0.184)
	[3.27]***	[1.649]*	[3.098]***	[0.484]	[7.104]***	[3.896]***	[1.332]	[0.745]
f_1	0.006**	0.016***	0.027***	-	0.018***	0.025***	0.013***	0.008*
f_2	0.019***	0.013**	0.016***	0.019***	-	0.009***	0.005**	-

Table A1.1: Benchmark Model Results (Continued…)

	Travel & Leisure	Fixed Line Telecom.	Banks	Non-life Insurance	Life Insurance	General Financial	Equity Inv. & Inst.	Soft. & Com Serv.
Panel B: Goodness-of-fit And Model Selection Criteria								
$\bar{R}^2$	0.460	0.171	0.559	0.339	0.706	0.680	0.431	0.314
AIC	-3.565	-2.035	-3.677	-3.200	-3.963	-3.736	-3.884	-2.436
BIC	-3.328	-1.797	-3.423	-2.979	-3.743	-3.499	-3.630	-2.199
Panel C: Model Diagnostics								
F-Statistic	23.415***	5.668***	32.469***	9.433***	67.047***	53.493***	18.779***	10.218***
JB	0.417	4.700	1.476	13.088***	1.721	1.194	0.352	10.098***
$Q(1)$	0.002	2.134	2.088	3.290*	3.996*	1.412	9.447***	0.002
$Q(5)$	1.722	15.740***	2.830	9.905*	5.399	5.269	14.347**	1.632
$Q^2(1)$	0.246	0.006	0.150	0.036	0.013	0.0811	0.901	0.157
$Q^2(5)$	5.227	7.203	5.389	2.941	1.682	7.135	4.592	3.042
ARCH(1)	0.239	0.006	0.147	0.035	0.013	0.079	0.882	0.152
ARCH(5)	0.950	1.488	1.236	0.534	0.314	1.754	0.851	0.628
Panel D: Residual Variance								
$\sigma^2_{\varepsilon i}$	0.001493	0.006652	0.001610	0.002098	0.000980	0.001221	0.001049	0.006682
Panel E: Conditional Variance Structure (ARCH/GARCH)								
ω	0.001	0.007***	8.52E-05	2.05E-03***	9.23E-04***	0.001***	0.0001	0.0001
α_1	0.265**	0.017	0.233**	0.018	0.052	0.093	0.055	0.101*
β_1	-	-	0.726***	-			0.793***	0.857***
F-Statistic	5.715**	0.058	82.118***	0.030	0.519	0.870	12.240***	493.679***
Panel F: Mean Errors								
ε_{it}	-0.0007593	0.0001324	0.0011064	-0.0000868	0.00017	-0.0001097	-0.000342	-0.0108835
Panel G: Theil's *U* And Decomposition								
Theil U	0.417	0.602	0.355	0.489	0.284	0.301	0.425	0.564
U_{BIAS}	0.000383	0.000003	0.000764	0.000004	0.000030	0.000010	0.000112	0.017506
U_{VAR}	0.182430	0.362134	0.085076	0.246571	0.075300	0.104240	0.162327	0.509499
U_{COV}	0.817187	0.637863	0.914160	0.753425	0.924671	0.895750	0.837561	0.472995

Table A1.2: Restricted Model Results

	Chemicals	Forestry & Paper	Ind. Met & Mining	Mining	Const & Materials	General Indust.	Elect. & Elect. Eq.	Industrial Eng.	Industrial Transport.
Panel A: Model Results									
Intercept	0.009	0.005	0.010	0.0003	0.002	0.009	0.001	0.010	0.006
	(0.003)	(0.005)	(0.009)	(0.006)	(0.005)	(0.003)	(0.003)	(0.004)	(0.004)
	[2.875]***	[0.969]	[1.137]	[0.057]	[0.526]	[2.877]***	[0.316]	[2.452]**	[1.624]
BP_{t-1}	0.034	-0.062	0.035	0.044	0.014	0.045	0.018	-0.006	0.033
	(0.036)	(0.061)	(0.094)	(0.052)	(0.04)	(0.028)	(0.028)	(0.037)	(0.042)
	[0.93]	[-1.027]	[0.368]	[0.861]	[0.337]	[1.576]	[0.617]	[-0.170]	[0.795]
$LEAD_{t-1}$	0.681	0.512	0.734	1.185	0.357	0.827	0.518	1.687	1.344
	(0.423)	(0.657)	(0.961)	(0.665)	(0.507)	(0.372)	(0.456)	(0.521)	(0.517)
	[1.608]	[0.78]	[0.764]	[1.781]*	[0.705]	[2.221]**	[1.136]	[3.239]***	[2.598]***
BUS_t	0.021	0.115	0.054	0.003	0.106	0.062	0.097	0.111	0.130
	(0.035)	(0.057)	(0.076)	(0.061)	(0.05)	(0.042)	(0.037)	(0.048)	(0.037)
	[0.611]	[2.005]**	[0.707]	[0.048]	[2.121]**	[1.49]	[2.644]***	[2.313]**	[3.493]***
$USD\varepsilon_t$	-0.180	0.243	-0.272	0.168	-0.511	-0.127	-0.309	-0.165	-0.213
	(0.11)	(0.178)	(0.366)	(0.187)	(0.172)	(0.116)	(0.115)	(0.148)	(0.128)
	[-1.647]*	[1.364]	[-0.742]	[0.898]	[-2.974]***	[-1.098]	[-2.696]***	[-1.116]	[-1.664]*
MET_t	0.165	0.467	0.619	0.42	0.185	0.003	0.327	0.444	0.11
	(0.074)	(0.131)	(0.23)	(0.125)	(0.086)	(0.063)	(0.089)	(0.123)	(0.099)
	[2.227]**	[3.556]***	[2.688]***	[3.357]***	[2.142]**	[0.049]	[3.658]***	[3.624]***	[1.108]
LTY_t	-1.718	-0.767	1.726	0.197	-5.129	-4.628	-4.263	-4.810	-7.827
	(1.053)	(1.716)	(3.344)	(1.765)	(1.885)	(1.208)	(1.032)	(1.546)	(1.356)
	[-1.633]*	[-0.447]	[0.516]	[0.112]	[-2.721]***	[-3.831]***	[-4.131]***	[-3.111]***	[-5.774]***
$TLI\varepsilon_t$	3.427	6.969	4.474	5.426	1.500	2.600	2.020	2.818	3.258
	(0.68)	(1.175)	(1.95)	(1.22)	(1.101)	(0.678)	(0.816)	(1.08)	(0.953)
	[5.04]***	[5.93]***	[2.295]**	[4.448]***	[1.363]	[3.834]***	[2.475]**	[2.608]***	[3.421]***

Table A1.2: Restricted Model Results (Continued…)

	Chemicals	Forestry & Paper	Ind. Met & Mining	Mining	Const & Materials	General Indust.	Elect. & Elect. Eq.	Industrial Eng.	Industrial Transport.
Panel B: Goodness-of-fit And Model Selection Criteria									
$\bar{R}^2$	0.144	0.120	0.095	0.167	0.145	0.164	0.225	0.238	0.212
AIC	-3.164	-2.070	-1.434	-2.298	-2.602	-3.281	-3.127	-2.734	-2.874
BIC	-2.977	-1.901	-1.248	-2.111	-2.432	-3.095	-2.958	-2.565	-2.688
Panel C: Model Diagnostics									
F-Statistic	8.197***	7.637***	2.621**	6.783***	6.448***	10.761***	10.093***	6.867***	14.495***
JB	6.623**	50.024***	100.745***	0.377	2.731	0.641	14.018***	125.614***	10.134***
$Q(1)$	3.209*	0.896	0.938	7.6160***	1.605	5.448**	0.002	0.025	5.514**
$Q(5)$	9.598*	9.768*	13.210**	15.515***	12.905**	9.373*	3.799	4.180	9.023
$Q^2(1)$	0.011	0.077	0.510	1.868	0.037	0.020	2.169	0.008	0.324
$Q^2(5)$	2.393	3.317	1.606	5.912	2.108	1.514	3.833	4.746	5.184
ARCH(1)	0.010	0.074	0.496	1.832	0.035	0.020	2.127	0.008	0.316
ARCH(5)	0.434	0.630	0.289	1.107	0.433	0.294	0.764	0.884	0.967
Panel D: Residual Variance									
$\sigma^2_{\varepsilon i}$	0.002496▲***	0.007075▲*	0.014494▲***	0.005324▲***	0.004002▲***	0.002072▲***	0.002251▲***	0.003448▲**	0.003230▲**
Panel E: Conditional Variance Structure (ARCH/GARCH)									
ω	0.0001	0.005***	8.17E-04	0.0007	0.003***	4.61E-05	0.002***	0.003***	0.001**
α_1	0.162*	0.273	0.123	0.035	0.138	0.068	0.172**	0.016	0.230**
β_1	0.789***		0.835***	0.823***		0.907***			0.419**
F-Statistic	92.128***	0.526	86.548***	7.282***	2.326	518.033***	4.266**	0.214	9.366***
Panel F: Mean Errors									
ε_{it}	-0.0027858	-0.0026095	-0.0022243	0.0005314	-0.0006423	-0.000255	-0.0001884	-0.0002243	-0.0018439
Panel G: Theil's *U* And Decomposition									
Theil U	0.633	0.640	0.706	0.624	0.646	0.604	0.573	0.553	0.565
U_{BIAS}	0.003115	0.000967	0.000343	0.000053	0.000104	0.000032	0.000016	0.000015	0.001057
U_{VAR}	0.436187	0.378917	0.543868	0.398490	0.431848	0.391294	0.324398	0.316895	0.287454
U_{COV}	0.560698	0.620117	0.455789	0.601456	0.568048	0.608674	0.675586	0.683090	0.711488
Panel H: Likelihood Ratio Test For Omitted Factors									
$M\varepsilon_t$	27.265***	50.273***	62.876***	216.132***	28.277***	78.702***	46.706***	26.603***	48.798***
$IM\varepsilon_t$	2.980*	1.197	3.402*	6.050**	2.392	0.169	7.910***	5.944**	1.443
f_{1t},f_{2t}	37.744***	5.846*	85.833***	90.541***	82.063***	42.183***	54.718***	36.945***	93.729***

Table A1.2: Restricted Model Results (Continued…)

	Support Services	Automobiles & Parts	Beverages	Food Producers	Health Care Eq. & Serv.	Pharm. & Biotech.	Food & Drug Retail.	General Retailers	Media
Panel A: Model Results									
Intercept	0.002	0.006	0.008	0.009	0.011	0.014	0.011	0.009	0.019
	(0.004)	(0.006)	(0.004)	(0.003)	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)
	[0.425]	[0.985]	[1.783]*	[3.223]***	[2.764]***	[2.992]***	[3.018]***	[2.146]**	[3.571]***
BP_{t-1}	0.045	-0.011	0.017	0.029	0.017	-0.019	0.132	0.077	0.036
	(0.039)	(0.052)	(0.035)	(0.025)	(0.047)	(0.042)	(0.037)	(0.038)	(0.053)
	[1.165]	[-0.218]	[0.484]	[1.185]	[0.353]	[-0.448]	[3.592]***	[2.053]**	[0.683]
$LEAD_{t-1}$	0.381	1.146	1.102	0.743	0.735	1.614	0.581	1.217	1.015
	(0.462)	(0.722)	(0.469)	(0.309)	(0.417)	(0.565)	(0.498)	(0.584)	(0.656)
	[0.823]	[1.588]	[2.352]**	[2.406]**	[1.764]*	[2.858]***	[1.166]	[2.085]**	[1.546]
BUS_t	0.119	0.087	0.042	0.074	0.078	0.075	0.160	0.084	0.161
	(0.052)	(0.074)	(0.051)	(0.036)	(0.044)	(0.056)	(0.042)	(0.053)	(0.07)
	[2.297]**	[1.183]	[0.824]	[2.049]**	[1.760]*	[1.324]	[3.843]***	[1.604]	[2.317]**
$USD\varepsilon_t$	-0.273	-0.026	0.079	-0.074	-0.186	-0.468	-0.283	-0.434	-0.056
	(0.136)	(0.193)	(0.16)	(0.095)	(0.14)	(0.173)	(0.147)	(0.128)	(0.187)
	[-2.001]*	[-0.137]	[0.494]	[-0.783]	[-1.328]	[-2.700]***	[-1.927]*	[-3.383]***	[-0.302]
MET_t	0.164	0.231	-0.069	0.055	0.001	0.063	-0.201	0.006	-0.014
	(0.08)	(0.16)	(0.092)	(0.058)	(0.075)	(0.096)	(0.083)	(0.092)	(0.116)
	[2.044]**	[1.446]	[-0.758]	[0.941]	[0.008]	[0.653]	[-2.412]**	[0.061]	[-0.12]
LTY_t	-3.369	-4.283	1.444	-5.335	-4.702	-6.654	-4.950	-8.799	-5.917
	(1.347)	(2.021)	(1.3)	(0.979)	(1.54)	(1.482)	(1.418)	(1.611)	(2.358)
	[-2.5]**	[-2.119]**	[1.111]	[-5.451]***	[-3.052]***	[-4.489]***	[-3.490]***	[-5.462]***	[-2.509]**
$TLI\varepsilon_t$	2.831	5.096	3.393	2.605	2.928	1.293	0.413	1.526	5.054
	(0.964)	(1.813)	(0.932)	(0.752)	(1.025)	(1.486)	(0.956)	(1.223)	(1.308)
	[2.938]***	[2.81]***	[3.64]***	[3.465]***	[2.858]***	[0.870]	[0.432]	[1.247]	[3.864]***

Table A1.2: Restricted Model Results (Continued…)

	Support Services	Automobiles & Parts	Beverages	Food Producers	Health Care Eq. & Serv.	Pharm. & Biotech.	Food & Drug Retail.	General Retailers	Media
Panel B: Goodness-of-fit And Model Selection Criteria									
$\bar{R}^2$	0.145	0.067	0.080	0.223	0.133	0.124	0.175	0.246	0.160
AIC	-3.042	-2.064	-2.803	-3.597	-2.972	-2.511	-3.005	-2.822	-2.298
BIC	-2.855	-1.877	-2.634	-3.411	-2.785	-2.325	-2.835	-2.652	-2.111
Panel C: Model Diagnostics									
F-Statistic	5.645***	2.768***	4.949***	7.421***	4.740***	7.656***	8.680***	13.082***	5.869***
JB	3.794	21.149***	5.338*	2.077	2.160	10.446***	0.669	15.442***	1.410
$Q(1)$	0.072	0.014	1.953	0.054	6.E-05	3.423*	0.085	0.017	0.380
$Q(5)$	5.813	1.601	2.828	4.183	8.313	6.128	1.609	4.750	9.514*
$Q^2(1)$	0.062	0.685	0.004	0.228	1.748	1.181	0.394	0.020	0.418
$Q^2(5)$	2.137	1.870	3.380	7.713	2.684	2.267	4.810	0.440	1.456
ARCH(1)	0.061	0.669	0.004	0.222	1.723	1.153	0.384	0.019	0.406
ARCH(5)	0.467	0.285	0.649	1.842	0.576	0.422	1.120	0.492	0.268
Panel D: Residual Variance									
$\sigma^2_{\varepsilon i}$	0.002576▲**	0.007268▲	0.003227▲***	0.001469▲***	0.002896▲***	0.004463▲	0.002663▲***	0.003169▲***	0.006042▲***
Panel E: Conditional Variance Structure (ARCH/GARCH)									
ω	0.0006	0.001**	0.003***	0.001***	0.000	0.001	0.002***	0.003***	0.0005**
α_1	0.113	0.035	-0.041	0.082	0.101	-0.042	0.171	0.041	0.035
α_2				0.134					
β_1	0.642**	0.826***			0.804***	0.760***			0.855***
F-Statistic	16.015***	140.359***	0.259	1.214	79.591***	10.540***	1.785	0.252	266.145***
Panel F: Mean Errors									
ε_{it}	-.0013621	-.0022257	.0004665	-.0005846	.0018479	-.0000708	-.0012425	-.0001647	-.0040142
Panel G: Theil's *U* And Decomposition									
Theil *U*	0.621	0.696	0.683	0.547	0.628	0.625	0.582	0.550	0.612
U_{BIAS}	0.000723	0.000685	0.000068	0.000234	0.001184	0.000001	0.000582	0.000009	0.002674
U_{VAR}	0.354891	0.461965	0.484036	0.305307	0.440557	0.407495	0.337370	0.310616	0.439903
U_{COV}	0.644386	0.537351	0.515897	0.694459	0.558259	0.592504	0.662047	0.689376	0.557423
Panel H: Likelihood Ratio Test For Omitted Factors									
$M\varepsilon_t$	41.503***	8.125***	72.389***	43.574***	27.480***	8.526***	13.851***	34.315***	54.962***
$IM\varepsilon_t$	1.817	0.222	0.044	0.338	1.394	0.299	0.123	3.178*	1.987
f_{1t}, f_{2t}	68.871***	21.734***	16.046***	53.326***	40.085***	29.474***	72.343***	151.977***	35.729***

Table A1.2: Restricted Model Results (Continued…)

	Travel & Leisure	Fixed Line Telecom.	Banks	Non-life Insurance	Life Insurance	General Financial	Equity Inv. & Inst.	Soft. & Com Serv.
				Panel A: Model Results				
Intercept	0.008	0.003	0.009	0.006	0.007	0.010	0.006	0.009
	(0.003)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.005)
	[2.427]**	[0.401]	[2.569]***	[1.512]	[1.688]*	[2.575]***	[1.927]*	[1.896]*
BP_{t-1}	-0.041	0.055	0.025	-0.015	0.000	0.051	0.061	0.025
	(0.03)	(0.065)	(0.042)	(0.035)	(0.035)	(0.034)	(0.033)	(0.043)
	[-1.339]	[0.858]	[0.591]	[-0.421]	[-0.013]	[1.464]	[1.834]*	[0.588]
$LEAD_{t-1}$	0.369	0.840	0.441	0.369	1.168	0.913	0.301	0.972
	(0.475)	(0.819)	(0.473)	(0.587)	(0.475)	(0.456)	(0.357)	(0.553)
	[0.776]	[1.025]	[0.932]	[0.629]	[2.459]**	[2.001]**	[0.843]	[1.758]*
BUS_t	0.046	0.020	0.099	0.070	0.054	0.058	0.064	0.050
	(0.04)	(0.065)	(0.045)	(0.04)	(0.043)	(0.043)	(0.037)	(0.052)
	[1.159]	[0.306]	[2.214]**	[1.753]*	[1.268]	[1.346]	[1.716]*	[0.954]
$USD\varepsilon_t$	-0.121	-0.054	-0.338	-0.192	-0.127	-0.289	0.060	-0.183
	(0.126)	(0.208)	(0.14)	(0.126)	(0.118)	(0.141)	(0.128)	(0.194)
	[-0.957]	[-0.261]	[-2.425]**	[-1.527]	[-1.078]	[-2.047]**	[0.472]	[-0.945]
MET_t	0.120	0.273	0.157	0.102	0.107	0.183	-0.006	0.156
	(0.083)	(0.123)	(0.091)	(0.097)	(0.072)	(0.075)	(0.074)	(0.089)
	[1.444]	[2.230]**	[1.717]*	[1.05]	[1.484]	[2.431]**	[-0.086]	[1.751]*
LTY_t	-5.063	-2.430	-5.196	-4.644	-5.931	-5.770	-0.995	-5.156
	(1.251)	(2.818)	(1.33)	(1.315)	(1.299)	(1.562)	(1.139)	(2.204)
	[-4.049]***	[-0.862]	[-3.906]***	[-3.532]***	[-4.566]***	[-3.693]***	[-0.874]	[-2.339]**
$TLI\varepsilon_t$	2.822	2.727	1.979	2.183	4.840	2.545	1.869	2.697
	(0.973)	(1.584)	(0.977)	(1.191)	(0.969)	(0.912)	(0.732)	(1.175)
	[2.901]***	[1.722]*	[2.025]**	[1.833]*	[4.995]***	[2.79]***	[2.553]**	[2.294]**

Table A1.2: Restricted Model Results (Continued…)

	Travel & Leisure	Fixed Line Telecom.	Banks	Non-life Insurance	Life Insurance	General Financial	Equity Inv. & Inst.	Soft. & Com Serv.
Panel B: Goodness-of-fit and Model Selection Criteria								
$\bar{R}^2$	0.071	0.032	0.108	0.101	0.255	0.162	0.049	0.039
AIC	-3.053	-1.900	-2.881	-2.920	-3.092	-2.951	-3.393	-2.242
BIC	-2.883	-1.731	-2.695	-2.750	-2.905	-2.765	-3.207	-2.055
Panel C: Model Diagnostics								
F-Statistic	5.904***	1.748*	6.938***	3.164***	11.348***	8.354***	3.299***	5.558***
JB	16.653***	4.849*	0.706	6.258**	1.327	0.686	4.754*	3.940
$Q(1)$	0.293	0.962	7.202***	0.735	3.910**	0.100	5.268**	0.175
$Q(5)$	2.805	8.907	11.080*	6.617	8.340	3.771	8.993	2.151
$Q^2(1)$	0.033	0.004	0.368	0.080	0.287	1.203	0.0002	0.311
$Q^2(5)$	1.405	6.044	6.015	8.010	4.221	6.247	0.306	2.096
ARCH(1)	0.032	0.004	0.359	0.078	0.280	1.178	0.0002	0.303
ARCH(5)	0.280	1.339	1.107	1.757	0.817	1.051	0.095	0.442
Panel D: Residual Variance								
$\sigma^2_{\varepsilon i}$	0.052	0.089	0.058	0.054				
Panel E: Conditional Variance Structure (ARCH/GARCH)								
ω	0.002***	0.008***	0.0003	0.003***	0.0001	0.0001	0.0004	0.0001
α_1	0.275**	0.022	0.192**	0.093	0.068	0.109**	0.063	0.088*
α_1								
β_1			0.703***		0.841***	0.679**	0.869***	
F-Statistic	4.078**	0.096	59.289***	0.824	180.338***	370.757***	6.695***	924.802***
Panel F: Mean Errors								
ε_{it}	-.0023466	.0002358	-.0036364	-.0004465	-.0048005	-.0065417	-.0009461	-.0140036
Panel G: Theil's *U* And Decomposition								
Theil *U*	0.663	0.767	0.643	0.678	0.549	0.610	0.707	0.761
U_{BIAS}	0.002078	0.000007	0.003990	0.000069	0.009177	0.013168	0.000502	0.020372
U_{VAR}	0.403605	0.591610	0.399462	0.479371	0.303879	0.378528	0.505566	0.613429
U_{COV}	0.594317	0.408383	0.596548	0.520560	0.686944	0.608304	0.493931	0.366200
Panel H: Likelihood Ratio Test For Omitted Factors								
$M\varepsilon_t$	72.389***	43.574***	27.480***	8.526***	95.889***	75.280***	63.818***	38.193***
$IM\varepsilon_t$	0.044	0.338	1.394	0.299	19.079***	9.625***	1.236	1.129
f_{1t}, f_{2t}	16.046***	53.326***	40.085***	29.474***	32.890***	45.549***	23.899***	3.380

Table A1.3: Unrestricted Market Model Results

	Chemicals	Forestry & Paper	Ind. Met & Mining	Mining	Const & Materials	General Indust.	Elect. & Elect. Eq.	Industrial Eng.	Industrial Transport.
					Panel A: Model Results				
Intercept	0.007	0.002	0.006	-0.0003	0.003	0.009	0.001	0.010	0.004
	(0.003)	(0.005)	(0.006)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)
	[2.371]**	[0.445]	[1.019]	[-0.105	[0.754]	[3.209]***	[0.27]	[2.599]***	[1.205]
BP_{t-1}	0.050	0.026	0.011	0.043	0.021	0.054	0.025	-0.003	0.076
	(0.031)	(0.048)	(0.058)	(0.031)	(0.039)	(0.024)	(0.034)	(0.03)	(0.038)
	[1.626]	[0.537]	[0.184]	[1.369]	[0.546]	[2.218]**	[0.74]	[-0.1]	[1.971]**
$LEAD_{t-1}$	0.832	0.721	1.390	1.454	0.240	0.997	0.536	1.704	1.450
	(0.424)	(0.66)	(0.971)	(0.335)	(0.515)	(0.262)	(0.381)	(0.498)	(0.435)
	[1.96]**	[1.092]	[1.431]	[4.346]***	[0.465]	[3.804]***	[1.407]	[3.424]***	[3.336]***
BUS_t	-0.006	0.132	-0.049	-0.014	0.099	0.067	0.091	0.110	0.123
	(0.034)	(0.06)	(0.062)	(0.036)	(0.05)	(0.035)	(0.039)	(0.044)	(0.031)
	[-0.162]	[2.186]**	[-0.797]	[-0.384]	[1.99]**	[1.902]*	[2.337]**	[2.479]**	[3.968]***
$USD\varepsilon_t$	-0.191	0.243	-0.276	0.269	-0.481	-0.168	-0.301	-0.167	-0.314
	(0.106)	(0.173)	(0.204)	(0.113)	(0.149)	(0.083)	(0.108)	(0.132)	(0.108)
	[-1.803]*	[1.406]	[-1.351]	[2.395]**	[-3.22]***	[-2.03]**	[-2.778]***	[-1.259]	[-2.9]***
MET_t	0.172	0.352	0.568	0.335	0.151	0.014	0.337	0.44	0.161
	(0.076)	(0.111)	(0.173)	(0.062)	(0.085)	(0.065)	(0.068)	(0.112)	(0.075)
	[2.268]**	[3.158]***	[3.278]***	[5.42]***	[1.765]*	[0.211]	[4.94]***	[3.912]***	[2.146]**
LTY_t	-1.435	-2.584	3.885	0.838	-4.939	-4.943	-4.020	-4.822	-7.629
	(1.086)	(1.595)	(2.144)	(0.942)	(1.57)	(0.834)	(1.205)	(1.377)	(1.227)
	[-1.321]	[-1.62]	[1.811]*	[0.889]	[-3.147]***	[-5.928]***	[-3.336]***	[-3.503]***	[-6.216]***
$TLI\varepsilon_t$	3.445	6.109	6.236	5.741	1.507	2.920	1.889	2.796	2.951
	(0.645)	(1.305)	(2.106)	(0.74)	(0.944)	(0.589)	(0.731)	(1.012)	(0.739)
	[5.344]***	[4.681]***	[2.961]***	[7.762]***	[1.596]	[4.958]***	[2.584]***	[2.763]***	[3.995]***
$M\varepsilon_t$	0.396	1.056	1.375	1.428	0.570	0.685	0.538	0.528	0.624
	(0.083)	(0.2)	(0.154)	(0.091)	(0.112)	(0.073)	(0.076)	(0.138)	(0.093)
	[4.746]***	[5.279]***	[8.915]***	[15.654]***	[5.09]***	[9.417]***	[7.09]***	[3.826]***	[6.742]***

Table A1.3: Unrestricted Market Model Results (Continued…)

	Chemicals	Forestry & Paper	Ind. Met & Mining	Mining	Const & Materials	General Indust.	Elect. & Elect. Eq.	Industrial Eng.	Industrial Transport.
Panel B: Goodness-of-fit And Model Selection Criteria									
$\bar{R}^2$	0.276	0.356	0.190	0.653	0.244	0.435	0.375	0.332	0.373
AIC	-3.295	-2.322	-1.751	-3.410	-2.739	-3.672	-3.371	-2.862	-3.094
BIC	-3.092	-2.135	-1.548	-3.207	-2.552	-3.486	-3.185	-2.676	-2.907
Panel C: Model Diagnostics									
F-Statistic	10.315***	13.709***	14.649***	53.629***	12.590***	36.334***	17.734***	7.640***	14.581***
JB	9.711***	149.860***	60.494***	2.453	3.666	4.094	1.466	40.524***	16.727***
$Q(1)$	2.468	0.229	0.008	0.400	1.723	2.058	0.227	0.005	2.170
$Q(5)$	8.533	9.777*	14.994***	2.490	9.086	5.052	8.669	5.131	4.869
$Q^2(1)$	0.033	0.013	0.302	2.589	0.005	0.050	0.002	0.095	0.790
$Q^2(5)$	4.787	3.654	2.817	5.317	1.158	1.748	2.929	3.812	3.855
ARCH(1)	0.032	0.013	0.293	2.547	0.004	0.049	0.002	0.092	0.772
ARCH(5)	0.864	0.629	0.510	1.074	0.190	0.357	0.538	0.733	0.681
Panel D: Residual Variance									
$\sigma^2_{\varepsilon i}$	0.002104▲*▼	0.005160▲▼	0.012903▲*▼	0.002205▲*▼*	0.003518▲*▼	0.001393▲*▼*	0.001805▲*▼	0.003005▲*▼	0.002558▲*▼
Panel E: Conditional Variance Structure (ARCH/GARCH)									
ω	0.0001	0.005***	0.0004	0.0001	0.003***	0.001***	0.002***	0.003***	0.002***
α_1	0.151**	-0.018	0.306***	0.144**	0.243**	0.167***	0.033	0.019	0.407***
β_1	0.795***		0.729***	0.842***					
F-Statistic	78.992***	2.386	104.732***	330.785***	4.403**	2.781*	0.092	0.034	9.575***
Panel F: Mean Errors									
ε_{it}	-0.0011691	0.0004707	0.001412	0.0011773	-0.001583	0.0000917	-0.0000166	-0.000154	0.0000231
Panel G: Theil's *U* And Decomposition									
Theil *U*	0.545	0.485	0.567	0.316	0.567	0.417	0.473	0.491	0.459
U_{BIAS}	0.000652	0.000043	0.000155	0.000632	0.000716	0.000006	0.000000	0.000008	0.000000
U_{VAR}	0.349547	0.236637	0.260387	0.092939	0.345242	0.132912	0.220679	0.248118	0.166944
U_{COV}	0.649801	0.763320	0.739458	0.906429	0.654043	0.867082	0.779321	0.751874	0.833056
Panel H: Likelihood Ratio Test For Omitted Factors									
$IM\varepsilon_t$	4.899**	6.076**	2.208	18.175***	3.416*	0.020	9.524***	7.403***	2.645
f_{1t}, f_{2t}	42.062***	2.143	94.798***	159.321***	85.181***	75.794***	66.008***	39.963***	129.552***

Table A1.3: Unrestricted Market Model Results (Continued…)

	Support Services	Automobiles & Parts	Beverages	Food Producers	Health Care Eq. & Serv.	Pharm. & Biotech.	Food & Drug Retail.	General Retailers	Media
Panel A: Model Results									
Intercept	0.002	0.005	0.010	0.009	0.011	0.014	0.011	0.009	0.017
	(0.003)	(0.006)	(0.004)	(0.003)	(0.004)	(0.005)	(0.003)	(0.004)	(0.005)
	[0.762]	[0.966]	[2.847]***	[3.421]***	[3.069]***	[3.071]***	[3.239]***	[2.313]**	[3.62]***
BP_{t-1}	0.048	-0.006	0.001	0.026	0.023	-0.020	0.133	0.087	0.034
	(0.034)	(0.054)	(0.04)	(0.024)	(0.044)	(0.043)	(0.036)	(0.032)	(0.054)
	[1.407]	[-0.112]	[0.019]	[1.1]	[0.532]	[-0.455]	[3.727]***	[2.764]***	[0.639]
$LEAD_{t-1}$	0.429	1.094	1.154	0.735	0.802	1.583	0.607	1.214	0.965
	(0.402)	(0.715)	(0.44)	(0.299)	(0.392)	(0.546)	(0.481)	(0.551)	(0.623)
	[1.067]	[1.529]	[2.624]***	[2.461]**	[2.045]**	[2.898]***	[1.261]	[2.203]**	[1.55]
BUS_t	0.103	0.079	0.056	0.078	0.066	0.078	0.153	0.072	0.152
	(0.044)	(0.075)	(0.04)	(0.029)	(0.04)	(0.055)	(0.039)	(0.05)	(0.063)
	[2.352]**	[1.057]	[1.387]	[2.667]***	[1.676]*	[1.42]	[3.864]***	[1.441]	[2.403]**
$USD\varepsilon_t$	-0.302	-0.038	0.032	-0.073	-0.157	-0.460	-0.286	-0.431	-0.016
	(0.109)	(0.182)	(0.103)	(0.089)	(0.138)	(0.171)	(0.129)	(0.126)	(0.158)
	[-2.767]***	[-0.208]	[0.309]	[-0.829]	[-1.134]	[-2.691]***	[-2.216]**	[-3.416]***	[-0.103]
MET_t	0.182	0.245	-0.091	0.072	0.003	0.061	-0.186	0.01	0.018
	(0.072)	(0.156)	(0.077)	(0.049)	(0.074)	(0.094)	(0.08)	(0.082)	(0.086)
	[2.525]**	[1.577]	[-1.185]	[1.466]	[0.042]	[0.654]	[-2.318]**	[0.127]	[0.205]
LTY_t	-3.243	-3.925	1.526	-4.914	-4.590	-6.609	-5.035	-8.659	-4.902
	(0.955)	(1.967)	(1.194)	(0.782)	(1.371)	(1.576)	(1.287)	(1.547)	(1.811)
	[-3.395]***	[-1.996]**	[1.278]	[-6.288]***	[-3.348]***	[-4.192]***	[-3.912]***	[-5.599]***	[-2.707]***
$TLI\varepsilon_t$	2.875	5.355	2.935	2.564	2.985	1.240	0.610	1.611	5.302
	(0.788)	(1.778)	(0.769)	(0.612)	(0.938)	(1.509)	(0.888)	(1.206)	(0.928)
	[3.647]***	[3.012]***	[3.818]***	[4.189]***	[3.181]***	[0.822]	[0.687]	[1.336]	[5.712]***
$M\varepsilon_t$	0.541	0.424	0.797	0.435	0.488	0.340	0.340	0.581	0.908
	(0.083)	(0.14)	(0.086)	(0.062)	(0.109)	(0.106)	(0.096)	(0.084)	(0.123)
	[6.517]***	[3.03]***	[9.267]***	[6.975]***	[4.482]***	[3.204]***	[3.536]***	[6.881]***	[7.405]***

Table A1.3: Unrestricted Market Model Results (Continued…)

	Support Services	Automobiles & Parts	Beverages	Food Producers	Health Care Eq. & Serv.	Pharm. & Biotech.	Food & Drug Retail.	General Retailers	Media
Panel B: Goodness-of-fit And Model Selection Criteria									
$\bar{R}^2$	0.281	0.113	0.362	0.381	0.245	0.161	0.219	0.369	0.344
AIC	-3.247	-2.096	-3.170	-3.809	-3.104	-2.545	-3.067	-2.990	-2.572
BIC	-3.044	-1.892	-2.983	-3.622	-2.901	-2.342	-2.880	-2.804	-2.369
Panel C: Model Diagnostics									
F-Statistic	13.446***	3.052***	15.231***	15.289***	8.564***	7.978***	9.956***	17.619***	16.268***
JB	2.271	23.291***	29.548***	4.022	2.982	18.809***	1.083	43.387***	3.831
$Q(1)$	0.329	0.035	1.478	0.101	0.208	3.305*	0.733	1.430	0.033
$Q(5)$	3.835	1.273	4.682	2.773	9.042	6.543	5.187	6.528	7.447
$Q^2(1)$	0.013	1.210	0.315	0.001	1.667	0.278	0.458	0.002	2.628
$Q^2(5)$	2.419	1.814	0.901	4.846	4.971	2.209	2.186	4.174	3.290
ARCH(1)	0.013	1.183	0.306	0.001	1.636	0.270	0.451	0.002	2.584
ARCH(5)	0.501	0.251	0.177	0.969	0.872	0.411	0.414	1.447	0.701
Panel D: Residual Variance									
$\hat{\sigma}^2_{\varepsilon i}$	0.002151▲*▼	0.006876▲▼	0.002223▲▼*	0.001165▲*▼	0.002508▲*▼	0.004248▲▼	0.002508▲*▼	0.002639▲*▼	0.004704▲*▼
Panel E: Conditional Variance Structure (ARCH/GARCH)									
ω	0.0004	0.001**	0.002***	0.001***	0.0003	0.001	0.002***	0.003***	0.0003**
α_1	0.187	0.047	0.189*	0.017	0.095	-0.031	0.203*	-0.009	0.045
β_1	0.630***	0.849***			0.773***	0.775***			0.877***
F-Statistic	22.411***	218.293***	3.499*	0.027	44.464***	11.535***	2.765*	0.039	384.693***
Panel F: Mean Errors									
ε_{it}	-0.0022043	-0.0019596	-0.0018321	0.0000056	0.0013323	-0.0001425	-0.0010266	0.0000974	-0.0023999
Panel G: Theil's *U* And Decomposition									
Theil *U*	0.517	0.653	0.473	0.456	0.543	0.596	0.547	0.471	0.490
U_{BIAS}	0.002265	0.000561	0.001516	0.000000	0.000711	0.000005	0.000422	0.000004	0.001229
U_{VAR}	0.228847	0.412601	0.235237	0.197865	0.332294	0.377488	0.285564	0.226270	0.289834
U_{COV}	0.768888	0.586838	0.763248	0.781721	0.666996	0.622507	0.714014	0.773727	0.708937
Panel H: Likelihood Ratio Test For Omitted Factors									
$IM\varepsilon_t$	3.147*	0.508	0.057	0.044	1.518	0.386	0.055	4.206**	2.957*
f_{1t}, f_{2t}	100.321***	22.549***	33.994***	72.207***	49.817***	32.956***	76.028***	194.029***	49.882***

Table A1.3: Unrestricted Market Model Results (Continued…)

	Travel & Leisure	Fixed Line Telecom.	Banks	Non-life Insurance	Life Insurance	General Financial	Equity Inv. & Inst.	Soft. & Com Serv.
Panel A: Model Results								
Intercept	0.008	0.003	0.006	0.006	0.006	0.008	0.006	0.008
	(0.003)	(0.007)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)
	[2.548]**	[0.423]	[2.118]**	[1.619]	[2.228]**	[2.636]***	[2.09]**	[1.872]*
BP_{t-1}	-0.007	0.060	0.064	-0.014	0.018	0.053	0.062	0.044
	(0.03)	(0.057)	(0.027)	(0.032)	(0.027)	(0.03)	(0.03)	(0.04)
	[-0.225]	[1.048]	[2.4]**	[-0.442]	[0.665]	[1.797]*	[2.097]**	[1.092]
$LEAD_{t-1}$	0.241	0.755	0.483	0.380	1.442	1.112	0.341	1.387
	(0.373)	(0.788)	(0.433)	(0.563)	(0.382)	(0.382)	(0.285)	(0.556)
	[0.647]	[0.958]	[1.114]	[0.675]	[3.775]****	[2.908]***	[1.196]	[2.493]**
BUS_t	0.048	0.019	0.092	0.069	0.049	0.052	0.051	0.058
	(0.036)	(0.085)	(0.039)	(0.037)	(0.029)	(0.033)	(0.029)	(0.054)
	[1.33]	[0.222]	[2.371]**	[1.839]*	[1.682]*	[1.59]	[1.755]*	[1.075]
$USD\varepsilon_t$	-0.162	-0.064	-0.323	-0.200	-0.126	-0.202	0.050	-0.105
	(0.114)	(0.221)	(0.097)	(0.118)	(0.112)	(0.105)	(0.106)	(0.155)
	[-1.42]	[-0.289]	[-3.325]***	[-1.702]*	[-1.13]	[-1.923]*	[0.471]	[-0.675]
MET_t	0.09	0.281	0.277	0.109	0.125	0.217	0.005	0.214
	(0.057)	(0.145)	(0.076)	(0.083)	(0.069)	(0.067)	(0.057)	(0.11)
	[1.591]	[1.943]*	[3.638]***	[1.307]	[1.811]*	[3.22]***	[0.086]	[1.947]*
LTY_t	-4.971	-2.127	-2.885	-4.453	-5.325	-4.475	-0.815	-4.052
	(1.085)	(1.61)	(0.846)	(1.328)	(0.848)	(1.12)	(0.959)	(1.515)
	[-4.581]***	[-1.321]	[-3.409]***	[-3.354]***	[-6.278]***	[-3.996]***	[-0.85]	[-2.675]***
$TLI\varepsilon_t$	2.205	2.881	2.137	2.168	4.949	3.503	1.914	3.512
	(0.794)	(1.49)	(0.698)	(1.134)	(0.694)	(0.701)	(0.605)	(1.38)
	[2.776]***	[1.933]*	[3.064]***	[1.911]*	[7.128]***	[4.999]***	[3.164]***	[2.545]**
$M\varepsilon_t$	0.580	0.658	0.677	0.462	0.746	0.737	0.587	0.761
	(0.081)	(0.171)	(0.082)	(0.088)	(0.075)	(0.084)	(0.071)	(0.119)
	[7.183]***	[3.843]***	[8.266]***	[5.226]***	[9.946]***	[8.754]***	[8.315]***	[6.414]***

Table A1.3: Unrestricted Market Model Results (Continued…)

	Travel & Leisure	Fixed Line Telecom.	Banks	Non-life Insurance	Life Insurance	General Financial	Equity Inv. & Inst.	Soft. & Com Serv.
Panel B: Goodness-of-fit and Model Selection Criteria								
$\bar{R}^2$	0.227	0.111	0.310	0.200	0.500	0.410	0.317	0.272
AIC	-3.259	-1.980	-3.242	-3.037	-3.493	-3.333	-3.715	-2.430
BIC	-3.073	-1.793	-3.022	-2.850	-3.290	-3.129	-3.512	-2.227
Panel C: Model Diagnostics								
F-Statistic	12.732***	4.916***	30.633***	5.281***	27.614***	24.224***	15.651***	10.069***
JB	6.845**	4.086	0.369	19.195***	4.709*	0.208	3.201	7.596**
$Q(1)$	0.038	1.931	5.223**	1.824	3.140*	2.497	3.806*	0.084
$Q(5)$	1.798	13.296**	9.044	8.905	8.721	4.506	7.045	2.237
$Q^2(1)$	0.013	0.0002	0.003	0.059	0.813	0.027	0.635	0.438
$Q^2(5)$	4.733	3.947	6.492	6.056	5.594	1.256	3.819	2.671
ARCH(1)	0.013	0.0002	0.003	0.058	0.793	0.026	0.620	0.426
ARCH(5)	0.893	0.782	1.365	1.129	0.949	0.167	0.731	0.578
Panel D: Residual Variance								
$\sigma^2_{\varepsilon i}$	$0.002199^{\blacktriangle * \blacktriangledown}$	$0.007246^{\blacktriangle \blacktriangledown}$	$0.002561^{\blacktriangle * \blacktriangledown *}$	$0.002566^{\blacktriangle \blacktriangledown}$	$0.001666^{\blacktriangle * \blacktriangledown *}$	$0.002267^{\blacktriangle * \blacktriangledown *}$	$0.001280^{\blacktriangle \blacktriangledown *}$	$0.007134^{\blacktriangle \blacktriangledown}$
Panel E: Conditional Variance Structure (ARCH/GARCH)								
ω	0.001***	0.007***	0.001	0.002***	0.000	0.000	0.0002	0.0001
α_1	0.375***	-0.004	-0.068	0.106	0.051	0.133**	0.040	0.081
α_2			0.291***					
β_1			0.499*		0.915***	0.829***	0.820***	0.876***
F-Statistic	7.173***	0.005	12.265***	1.606	437.067***	556.717***	15.022***	750.565***
Panel F: Mean Errors								
ε_{it}	-0.0020737	-0.0000620	-0.0004952	-0.0004130	-0.0042658	-0.0042475	-0.0004180	-0.0126528
Panel G: Theil's *U* And Decomposition								
Theil *U*	0.544	0.665	0.502	0.588	0.395	0.452	0.495	0.589
U_{BIAS}	0.001962	0.000001	0.000096	0.000067	0.010862	0.007935	0.000137	0.022061
U_{VAR}	0.259756	0.442132	0.230788	0.360262	0.136530	0.213981	0.225104	0.500836
U_{COV}	0.738282	0.557867	0.769115	0.639671	0.852607	0.778083	0.774759	0.477103
Panel H: Likelihood Ratio Test For Omitted Factors								
$IM\varepsilon_t$	9.576***	3.591*	1.968	0.023	28.977***	13.864***	2.053	0.632
f_{1t}, f_{2t}	53.326***	13.102***	82.031***	38.299***	55.712***	68.961***	36.397***	5.970*

Table A1.4: Unrestricted Model Results

	Chemicals	Forestry & Paper	Ind. Met & Mining	Mining	Const & Materials	General Indust.	Elect. & Elect. Eq.	Industrial Eng.	Industrial Transport.
Panel A: Model Results									
Intercept	0.007	0.003	0.006	0.001	0.003	0.009	0.001	0.010	0.004
	(0.003)	(0.005)	(0.006)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)
	[2.323]**	[0.49]	[1.034]	[0.344]	[0.789]	[3.2]***	[0.263]	[2.646]***	[1.172]
BP_{t-1}	0.055	0.023	0.017	0.035	0.018	0.054	0.027	-0.003	0.070
	(0.031)	(0.046)	(0.06)	(0.028)	(0.038)	(0.024)	(0.033)	(0.03)	(0.037)
	[1.777]*	[0.505]	[0.277]	[1.241]	[0.475]	[2.196]**	[0.83]	[-0.098]	[1.896]*
$LEAD_{t-1}$	0.851	0.740	1.365	1.465	0.197	0.999	0.536	1.725	1.539
	(0.432)	(0.654)	(0.998)	(0.311)	(0.501)	(0.263)	(0.375)	(0.488)	(0.438)
	[1.969]**	[1.131]	[1.368]	[4.718]***	[0.393]	[3.805]***	[1.432]	[3.53]***	[3.516]***
BUS_t	-0.003	0.131	-0.042	-0.016	0.102	0.067	0.091	0.109	0.125
	(0.034)	(0.058)	(0.059)	(0.032)	(0.047)	(0.035)	(0.039)	(0.043)	(0.031)
	[-0.077]	[2.24]**	[-0.707]	[-0.513]	[2.193]**	[1.899]*	[2.303]**	[2.508]**	[3.985]***
$USD\varepsilon_t$	-0.216	0.246	-0.312	0.284	-0.482	-0.168	-0.294	-0.171	-0.300
	(0.107)	(0.168)	(0.226)	(0.109)	(0.148)	(0.083)	(0.105)	(0.128)	(0.108)
	[-2.027]**	[1.464]	[-1.378]	[2.594]***	[-3.266]***	[-2.019]**	[-2.797]***	[-1.338]	[-2.779]***
MET_t	0.169	0.353	0.588	0.33	0.146	0.014	0.333	0.437	0.148
	(0.078)	(0.108)	(0.183)	(0.058)	(0.084)	(0.066)	(0.067)	(0.106)	(0.075)
	[2.158]**	[3.275]***	[3.224]***	[5.644]***	[1.738]*	[0.211]	[4.963]***	[4.128]***	[1.97]**
LTY_t	-1.728	-2.573	3.676	0.449	-5.002	-4.942	-4.034	-4.845	-7.638
	(1.07)	(1.603)	(2.075)	(0.942)	(1.546)	(0.834)	(1.207)	(1.365)	(1.214)
	[-1.616]	[-1.605]	[1.772]*	[0.477]	[-3.236]***	[-5.923]***	[-3.343]***	[-3.549]***	[-6.29]***
$TLI\varepsilon_t$	3.232	6.124	5.868	6.052	1.387	2.914	1.911	2.770	2.820
	(0.659)	(1.262)	(1.882)	(0.74)	(0.959)	(0.599)	(0.756)	(0.954)	(0.76)
	[4.905]***	[4.854]***	[3.117]***	[8.18]***	[1.446]	[4.869]***	[2.526]**	[2.904]***	[3.708]***
$M\varepsilon_t$	0.413	1.053	1.355	1.424	0.577	0.685	0.543	0.535	0.634
	(0.086)	(0.193)	(0.15)	(0.084)	(0.113)	(0.073)	(0.075)	(0.133)	(0.09)
	[4.805]***	[5.458]***	[9.02]***	[16.979]***	[5.119]***	[9.328]***	[7.224]***	[4.025]***	[7.039]***
$IM\varepsilon_t$	0.244	0.457	0.317	-0.408	0.269	0.013	0.331	0.385	0.197
	(0.098)	(0.201)	(0.172)	(0.113)	(0.133)	(0.114)	(0.128)	(0.148)	(0.111)
	[2.488]**	[2.271]***	[1.845]*	[-3.612]***	[2.028]**	[0.115]	[2.584]***	[2.597]***	[1.775]*

Table A1.4: Unrestricted Model Results (Continued…)

	Chemicals	Forestry & Paper	Ind. Met & Mining	Mining	Const & Materials	General Indust.	Elect. & Elect. Eq.	Industrial Eng.	Industrial Transport.
				Panel B: Goodness-of-fit And Model Selection Criteria					
$\bar{R}^2$	0.304	0.373	0.195	0.678	0.250	0.432	0.400	0.352	0.385
AIC	-3.310	-2.343	-1.753	-3.494	-2.746	-3.662	-3.410	-2.890	-3.097
BIC	-3.090	-2.140	-1.532	-3.274	-2.542	-3.458	-3.207	-2.687	-2.894
				Panel C: Model Diagnostics					
F-Statistic	10.115***	12.538***	14.435***	57.984***	11.526***	32.073***	18.244***	7.516***	13.490***
JB	9.710*	111.2067***	58.579***	1.197	2.707	4.218	2.100	29.037***	18.655***
$Q(1)$	2.394	0.246	0.125	0.697	2.056	2.107	0.048	0.001	1.784
$Q(5)$	7.542	7.798	15.679***	5.017	11.593**	5.086	8.070	5.860	4.637
$Q^2(1)$	0.002	0.029	0.385	2.428	0.002	0.053	0.018	0.246	0.749
$Q^2(5)$	5.249	2.143	2.946	8.339	2.063	1.699	2.812	3.804	3.531
ARCH(1)	0.002	0.028	0.374	2.388	0.002	0.052	0.018	0.239	0.733
ARCH(5)	0.948	0.384	0.524	1.505	0.365	0.345	0.533	0.728	0.649
				Panel D: Residual Variance					
$\sigma^2_{\varepsilon i}$	0.002011$^{\blacktriangle\blacktriangledown}$	0.004997$^{-\blacktriangledown *}$	0.012748$^{\blacktriangle *\blacktriangledown}$	0.002033$^{\blacktriangle *\blacktriangledown *}$	0.003472$^{\blacktriangle *\blacktriangledown}$	0.001392$^{\blacktriangle *\blacktriangledown *}$	0.001724$^{\blacktriangle *\blacktriangledown *}$	0.002900$^{\blacktriangle\blacktriangledown}$	0.002499$^{\blacktriangle *\blacktriangledown}$
				Panel E: Conditional Variance Structure (ARCH/GARCH)					
ω	1.36E-04	0.005***	3.47E-04	3.12E-05	0.003***	0.001***	0.002***	0.003***	1.65E-03***
α_1	0.131**	-0.015	0.288***	0.123*	0.271**	0.166*	0.062	0.034	0.372***
β_1	0.809***		0.744***	0.875***	-	-	-		
F-Statistic	70.235***	0.484	118.800***	333.030***	4.727**	2.734*	0.321	0.090	9.193***
				Panel F: Mean Errors					
ε_{it}	-0.0010447	0.0002828	0.0015428	-0.0001554	-0.0016749	0.0001050	0.0000339	-0.0001607	0.0002403
				Panel G: Theil's *U* And Decomposition					
Theil *U*	0.526	0.473342	0.564082	0.301	0.558	0.417	0.456	0.479	0.453
U_{BIAS}	0.000545	0.000016	0.000188	0.000012	0.000812	0.000008	0.000001	0.000009	0.000023
U_{VAR}	0.334508	0.225970	0.266015	0.085444	0.331698	0.133269	0.204136	0.230278	0.168273
U_{COV}	0.664947	0.774013	0.733797	0.914544	0.667490	0.866723	0.795863	0.769713	0.831704
				Panel H: Likelihood Ratio Test For Omitted Factors					
f_{1t}, f_{2t}	48.057***	2.297	94.441***	328.169***	86.660***	76.166***	69.702***	42.817***	135.589***

Table A1.4: Unrestricted Model Results (Continued…)

	Support Services	Automobiles & Parts	Beverages	Food Producers	Health Care Eq. & Serv.	Pharm. & Biotech.	Food & Drug Retail.	General Retailers	Media
Panel A: Model Results									
Intercept	0.002	0.005	0.010	0.009	0.012	0.014	0.011	0.009	0.017
	(0.003)	(0.006)	(0.004)	(0.003)	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)
	[0.773]	[0.937]	[2.868]***	[3.42]***	[3.114]***	[3.056]***	[3.106]***	[2.339]**	[3.588]***
BP_{t-1}	0.057	-0.006	0.000	0.027	0.025	-0.020	0.133	0.087	0.035
	(0.035)	(0.054)	(0.04)	(0.024)	(0.043)	(0.043)	(0.036)	(0.032)	(0.053)
	[1.634]*	[-0.107]	[0.01]	[1.111]	[0.567]	[-0.456]	[3.73]***	[2.743]***	[0.675]
$LEAD_{t-1}$	0.493	1.065	1.153	0.739	0.816	1.563	0.609	1.215	0.914
	(0.393)	(0.71)	(0.439)	(0.299)	(0.397)	(0.548)	(0.485)	(0.541)	(0.615)
	[1.255]	[1.499]	[2.625]***	[2.471]**	[2.055]**	[2.85]***	[1.256]	[2.245]**	[1.486]
BUS_t	0.102	0.076	0.056	0.078	0.066	0.081	0.152	0.072	0.150
	(0.042)	(0.074)	(0.04)	(0.029)	(0.039)	(0.055)	(0.04)	(0.049)	(0.062)
	[2.395]**	[1.028]	[1.405]	[2.657]***	[1.679]*	[1.463]	[3.845]***	[1.468]	[2.418]**
$USD\varepsilon_t$	-0.304	-0.043	0.031	-0.073	-0.157	-0.459	-0.287	-0.430	-0.020
	(0.109)	(0.182)	(0.104)	(0.089)	(0.138)	(0.172)	(0.13)	(0.125)	(0.155)
	[-2.79]***	[-0.239]	[0.296]	[-0.826]	[-1.143]	[-2.665]***	[-2.209]**	[-3.442]***	[-0.13]
MET_t	0.168	0.25	-0.094	0.072	0.003	0.062	-0.187	0.011	0.014
	(0.071)	(0.154)	(0.077)	(0.049)	(0.074)	(0.094)	(0.082)	(0.084)	(0.083)
	[2.385]**	[1.626]*	[-1.225]	[1.456]	[0.047]	[0.659]	[-2.289]**	[0.128]	[0.167]
LTY_t	-3.339	-3.897	1.527	-4.914	-4.489	-6.580	-5.008	-8.655	-4.723
	(0.966)	(1.942)	(1.194)	(0.779)	(1.348)	(1.583)	(1.291)	(1.473)	(1.856)
	[-3.457]***	[-2.006]**	[1.278]	[-6.304]***	[-3.33]***	[-4.158]***	[-3.878]***	[-5.877]***	[-2.544]**
$TLI\varepsilon_t$	2.736	5.439	2.915	2.572	3.063	1.222	0.612	1.612	5.333
	(0.774)	(1.776)	(0.765)	(0.609)	(0.915)	(1.491)	(0.891)	(1.156)	(0.926)
	[3.536]***	[3.063]***	[3.811]***	[4.222]***	[3.347]***	[0.82]	[0.686]	[1.395]	[5.756]***
$M\varepsilon_t$	0.548	0.436	0.798	0.434	0.484	0.343	0.340	0.580	0.907
	(0.084)	(0.14)	(0.085)	(0.063)	(0.111)	(0.107)	(0.098)	(0.084)	(0.118)
	[6.534]***	[3.111]***	[9.347]***	[6.902]***	[4.367]***	[3.193]***	[3.483]***	[6.95]***	[7.662]***
$IM\varepsilon_t$	0.195	0.158	-0.029	-0.019	-0.144	0.108	0.030	0.273	0.277
	(0.121)	(0.195)	(0.121)	(0.099)	(0.141)	(0.172)	(0.146)	(0.163)	(0.168)
	[1.618]	[0.812]	[-0.241]	[-0.194]	[-1.023]	[0.628]	[0.206]	[1.678]*	[1.651]*

Table A1.4: Unrestricted Model Results (Continued…)

	Support Services	Automobiles & Parts	Beverages	Food Producers	Health Care Eq. & Serv.	Pharm. & Biotech.	Food & Drug Retail.	General Retailers	Media
Panel B: Goodness-of-fit And Model Selection Criteria									
$\bar{R}^2$	0.293	0.114	0.358	0.377	0.241	0.161	0.216	0.379	0.351
AIC	-3.253	-2.088	-3.160	-3.798	-3.102	-2.537	-3.057	-3.002	-2.578
BIC	-3.033	-1.867	-2.956	-3.595	-2.881	-2.317	-2.853	-2.798	-2.358
Panel C: Model Diagnostics									
F-Statistic	13.012***	2.802***	14.262***	13.595***	7.804***	7.293***	8.856***	17.519***	15.581***
JB	1.628	22.174***	28.143***	4.090	3.504	19.079***	1.238	23.747***	2.682
$Q(1)$	0.322	0.042	1.468	0.098	0.240	2.972	0.722	1.596	0.027
$Q(5)$	3.928	1.160	4.724	2.730	9.260*	6.112	5.057	6.787	6.798
$Q^2(1)$	0.023	1.179	0.334	0.001	1.406	0.215	0.454	0.002	2.339
$Q^2(5)$	2.272	1.839	0.896	4.933	3.752	1.960	2.293	3.742	2.932
ARCH(1)	0.023	1.153	0.325	0.001	1.378	0.209	0.447	0.002	2.296
ARCH(5)	0.461	0.245	0.176	0.981	0.669	0.360	0.420	1.128	0.621
Panel D: Residual Variance									
$\sigma^2_{\varepsilon i}$	0.002106▲*▼	0.006826▲▼	0.002226▲▼*	0.001164▲*▼*	0.002509▲*▼	0.004224▲▼	0.002505▲*▼	0.002582▲*▼	0.004625▲▼*
Panel E: Conditional Variance Structure (ARCH/GARCH)									
ω	0.0004	0.0005	0.002***	0.001***	0.0003	0.001	0.002***	0.003***	0.0002**
α_1	0.169	0.052	0.204*	0.020	0.105	-0.024	0.202	-0.011	0.043
β_1	0.662***	0.853***	-		0.770***	0.777***		-	0.880***
F-Statistic	25.061***	226.731***	3.703*	0.038	50.878***	11.108***	2.563	0.058	391.6122***
Panel F: Mean Errors									
ε_{it}	-0.0022074	-0.0017722	-0.0019114	0.0000036	0.0011776	-0.0001329	-0.0009629	0.0001003	-0.0021807
Panel G: Theil's *U* And Decomposition									
Theil U	0.508	0.647	0.473	0.456	0.541	0.594	0.547	0.463	0.485
U_{BIAS}	0.002321	0.000462	0.001647	0.000000	0.000555	0.000004	0.000372	0.000004	0.001033
U_{VAR}	0.222877	0.405146	0.234483	0.217582	0.325692	0.377666	0.286882	0.218932	0.288951
U_{COV}	0.774803	0.594392	0.763870	0.782418	0.673753	0.622330	0.712746	0.781064	0.710016
Panel H: Likelihood Ratio Test For Omitted Factors									
f_{1t}, f_{2t}	107.996***	22.873***	31.826***	72.180***	48.584***	35.353***	76.746***	201.931***	51.555***

Table A1.4: Unrestricted Model Results (Continued…)

	Travel & Leisure	Fixed Line Telecom.	Banks	Non-life Insurance	Life Insurance	General Financial	Equity Inv. & Inst.	Soft. & Com Serv.
Panel A: Model Results								
Intercept	0.006	0.003	0.005	0.006	0.003	0.007	0.005	0.008
	(0.003)	(0.006)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)
	[2.145]**	[0.542]	[1.846]*	[1.614]	[1.129]	[2.517]**	[1.873]*	[1.789]*
BP_{t-1}	-0.015	0.063	0.068	-0.014	0.019	0.059	0.060	0.044
	(0.029)	(0.067)	(0.028)	(0.032)	(0.023)	(0.028)	(0.028)	(0.04)
	[-0.516]	[0.947]	[2.431]**	[-0.438]	[0.799]	[2.142]**	[2.142]**	[1.109]
$LEAD_{t-1}$	0.313	0.744	0.579	0.384	1.389	0.911	0.279	1.362
	(0.378)	(0.742)	(0.45)	(0.562)	(0.358)	(0.369)	(0.294)	(0.596)
	[0.828]	[1.002]	[1.286]	[0.683]	[3.882]***	[2.466]**	[0.947]	[2.288]**
BUS_t	0.054	0.020	0.085	0.069	0.069	0.052	0.056	0.058
	(0.036)	(0.06)	(0.042)	(0.037)	(0.03)	(0.033)	(0.028)	(0.055)
	[1.532]	[0.333]	[2.017]**	[1.844]*	[2.316]**	[1.572]	[1.975]**	[1.064]
$USD\varepsilon_t$	-0.144	-0.073	-0.310	-0.202	-0.180	-0.228	0.062	-0.105
	(0.112)	(0.224)	(0.1)	(0.118)	(0.093)	(0.102)	(0.095)	(0.153)
	[-1.282]	[-0.328]	[-3.083]***	[-1.715]*	[-1.942]*	[-2.242]**	[0.649]	[-0.688]
MET_t	0.083	0.286	0.26	0.109	0.111	0.211	0.016	0.215
	(0.056)	(0.139)	(0.08)	(0.083)	(0.06)	(0.068)	(0.053)	(0.108)
	[1.48]	[2.055]**	[3.251]***	[1.312]	[1.856]*	[3.116]***	[0.309]	[1.98]**
LTY_t	-4.683	-1.956	-3.253	-4.449	-5.569	-4.370	-0.356	-3.971
	(1.08)	(2.963)	(0.87)	(1.325)	(0.927)	(1.055)	(0.862)	(1.499)
	[-4.337]***	[-0.66]	[-3.738]***	[-3.359]***	[-6.005]***	[-4.143]***	[-0.412]	[-2.649]***
$TLI\varepsilon_t$	2.242	2.963	2.087	2.163	4.757	3.331	1.576	3.474
	(0.814)	(1.524)	(0.675)	(1.135)	(0.637)	(0.691)	(0.533)	(1.351)
	[2.754]***	[1.944]*	[3.093]***	[1.907]*	[7.462]***	[4.822]***	[2.957]***	[2.573]***
$M\varepsilon_t$	0.583	0.669	0.695	0.463	0.732	0.717	0.564	0.756
	(0.077)	(0.152)	(0.083)	(0.089)	(0.063)	(0.079)	(0.067)	(0.121)
	[7.541]***	[4.413]***	[8.326]***	[5.211]***	[11.556]***	[9.105]***	[8.452]***	[6.244]***
$IM\varepsilon_t$	0.336	0.420	0.209	0.020	0.604	0.378	0.111	0.134
	(0.093)	(0.261)	(0.122)	(0.124)	(0.103)	(0.107)	(0.091)	(0.194)
	[3.61]***	[1.611]	[1.71]*	[0.165]	[5.88]***	[3.535]***	[1.228]	[0.687]

Table A1.4: Unrestricted Model Results (Continued…)

	Travel & Leisure	Fixed Line Telecom.	Banks	Non-life Insurance	Life Insurance	General Financial	Equity Inv. & Inst.	Soft. & Com Serv.
Panel B: Goodness-of-fit And Model Selection Criteria								
$\bar{R}^2$	0.260	0.121	0.330	0.196	0.585	0.434	0.321	0.276
AIC	-3.299	-1.988	-3.246	-3.027	-3.647	-3.395	-3.715	-2.423
BIC	-3.095	-1.785	-3.008	-2.823	-3.443	-3.174	-3.511	-2.203
Panel C: Model Diagnostics								
F-Statistic	13.676***	3.676***	23.894***	4.735***	34.003***	21.949***	13.781***	9.424***
JB	9.273***	4.958*	0.747	18.918***	7.375**	0.243	2.767	7.980**
$Q(1)$	0.413	2.174	5.300**	1.796	3.098	3.668*	2.864*	0.112
$Q(5)$	1.648	12.559**	10.300*	8.937	6.035	6.344	6.040	2.253
$Q^2(1)$	0.0003	0.004	0.003	0.061	0.108	0.130	0.017	0.380
$Q^2(5)$	3.671	5.670	8.587	6.091	5.766	0.864	5.195	2.772
ARCH(1)	0.0003	0.004	0.003	0.059	0.105	0.127	0.016	0.370
ARCH(5)	0.670	1.069	1.826	1.132	1.055	0.159	0.945	0.611
Panel D: Residual Variance								
$\hat{\sigma}^2_{\varepsilon i}$	0.001389$^{\blacktriangle\blacktriangledown *}$	0.002168$^{\blacktriangle * \blacktriangledown *}$	0.001266$^{\blacktriangle\blacktriangledown *}$	0.007063$^{\blacktriangle * \blacktriangledown}$				
Panel E: Conditional Variance Structure (ARCH/GARCH)								
ω	0.001***	7.28E-03***	6.05E-04	0.002***	0.001***	5.74E-05	0.001***	0.0001
α_1	0.371***	-0.028	-0.067	0.105	0.170*	0.160**	-0.023	0.080
α_2			0.262***					
β_1	-		0.507*			0.808***		0.878***
F-Statistic	8.332***	0.190	12.765***	1.561	2.758*	451.809***	0.224	768.383
Panel F: Mean Errors								
ε_{it}	-0.0007120	-0.0005689	-0.0000057	-0.0003988	-0.0010707	-0.0036239	0.000037	-0.0123881
Panel G: Theil's U And Decomposition								
Theil U	0.526	0.647	0.490	0.588	0.350	0.441	0.501	0.587
U_{BIAS}	0.000243	0.000046	0.000000	0.000062	0.000829	0.006053	0.000001	0.021373
U_{VAR}	0.248140	0.411973	0.222863	0.359678	0.108306	0.221824	0.260930	0.508700
U_{COV}	0.751617	0.587981	0.777137	0.640260	0.890865	0.772122	0.739069	0.469926
Panel H: Likelihood Ratio Test For Omitted Factors								
f_{1t}, f_{2t}	55.114***	12.948***	86.892***	38.431***	71.840***	75.515***	37.198***	6.100**