

**Financial time series
with heteroscedastic volatility
in the
South African financial markets**

MG Ferreira

Presented in partial fulfilment of the degree MSc
in the Faculty of Science
University of Pretoria
October 1999

SOLI DEO GLORIA

Preface

Time Series Analysis and the News Impact Curve

I was introduced to the subject of time series analysis while studying econometrics – at which time most of its approaches did not make sense to me. I was convinced that regression analysis held the answer to all my econometric headaches and viewed tools such as ARMA models with suspicion and as belonging to the study of the physical sciences.

Since econometrics is mostly concerned with the structural relationships between econometric variables this is not such an isolated experience. Practitioners often view these tools as not of crucial importance as they do not assist in spelling out e.g. the transmission mechanism from a change in interest rates to changes in inflation.

This attitude is not shared by those operating in the financial markets. Here the properties of a time series viewed in isolation is often of crucial importance. As the frequency of observations increases structural relationships become less meaningful and also less important. In the short term time series analysis then becomes a powerful analysis tool that can shed important light on a number of questions burning in the minds of traders, portfolio and risk managers.

Again my experience was not different. As I became involved in matters of finance so my interest in time series analysis grew. My own trading experience convinced me that the impact of ‘bad’ news was in general more violent than the impact of ‘good’ news and when I first encountered the *News Impact Curve* in theory it made a lot of sense to me as a tool of great practical value.

Constructing this curve however requires knowledge of ARCH models which in turns requires knowledge of ARMA models. In order to lay the

theoretical foundation necessary to investigate ARMA models we have to look at linear combinations of stationary time series as well as to difference equations. The best part of the first chapter is devoted to this investigation in our pilgrimage towards the news impact curve.

Fourier Series

We only make one detour in this journey – the section on linear cyclical models. This section is in many ways different to – even detached from – the rest of the chapter and as such presented many problems. It requires a different theoretical framework and the most important theorems in it¹ are stated without proof. Still the breath of fresh – or different – air that it provided during this study more than made up for the additional effort it required and the headaches it presented.

Difference Equations

In laying the foundation of linear time series models as we approach ARMA models one of the difficulties that presented itself was that of addressing difference equations. This portion – in line with so many other texts – initially contained very little theory of any mathematical significance. It left much to be desired and presented a serious challenge to me.

This changed drastically when I obtained Hamilton's book on time series analysis. In it he takes the necessary time to develop the theoretical tools needed later without excuse. I undertook this study in part because I never could find a text that dealt adequately and systematically with this theory in the context of time series analysis. Hamilton's book is in many ways that textbook and I am sure it will soon become very popular with many academic institutions.

One important tool that Hamilton did not use in his text is that of Vandermonde's matrices. He does invert the $\mathbf{F}$ matrix associated with a time series difference equation but without proof or mentioning that its eigenvectors form a Vandermonde's matrix. I needed a complete proof and

¹Those dealing with the Fourier series representation of a function.

also realized that this matrix had a structure to it that may have been studied already – though in a different context.

After consulting many texts on linear algebra I made what to me was the most rewarding ‘discovery’ of the whole study – that these eigenvectors form a *Vandermonde’s* matrix which can then easily be inverted. This discovery was made in Knuth’s well known *Fundamental Algorithms* where he briefly mentions Vandermonde’s as well as Cauchy’s and the combinatorial matrix as part of exercises on matrices and determinants!

The ARMA Moments

Once the theoretical framework is in place the treatment of ARMA models and even ARCH models is not that difficult. One exception lies in computing the moments of ARMA models. This section was added after completing the chapter on ARCH models where I calculate the ARCH moments after Engle (1982) and the GARCH moments after Teräsvirta (1996).

I could find no reference giving either the fourth moment or kurtosis of an ARMA model and after a long struggle obtained it as the same as that of the normal distribution. This is actually simple to show as I did in the text. The initial work was not lost as I used that to derive the kurtosis of the ARMA(1,1) model without having to assume normal errors.

Heteroscedastic Models

The past few years saw an explosion in ARCH type heteroscedastic time series models in finance. In the chapter dealing with these models I concentrate on the basic ARCH theory and do nothing more than stating the alternative representations. The two texts by Teräsvirta proved to be extremely helpful in writing this chapter and are worth obtaining as they contain a number of important and new results.

Empirical Study

Although the aim of this study was to develop the theoretical tools needed to construct the news impact curve the emphasis in the final chapter is more on the practical application of these tools than on obtaining the correct

results. Therefore not a lot of time is spent on ensuring that the resulting properties of the models make sense – the focus is more on obtaining these results than on the results themselves.

Still the study sheds some light on what has been taking place in South Africa in recent years as well as how this has impacted on volatility.

Notation

I have tried to follow a standardized notation throughout, but also in this did I make some mistakes. For subscripts I make use as far as possible of the letter i but this presented a problem when I encountered complex numbers. To overcome this $\iota = \sqrt{-1}$ is used to represent complex numbers and i remains an index variable. In a time series context y_t generally refers to a time series and $z_t = y_t - \mu_{y_t}$ refers to a zero mean time series. Finally I use $\mathbf{I}$ to indicate the information matrix and $\mathcal{I}$ as the identity matrix. Note that there is a short subject and notation index that will assist in reading this text.

Thanks

This study is in many ways the result of the effort of many people. I wish to specifically thank my supervisor, Prof FD van Niekerk, who allowed and assisted me to conduct such a study as part of a mathematics degree. He also insisted that I complete this work and hand it in – otherwise I would still be working it.

My wife, Marietjie, and daughter, Bianca, also deserves thanks for allowing me the time and assisting me to complete this study. My parents, MG and Ragie, have always encouraged me to seek after wisdom and knowledge and they also deserve thanks.

**Financial time series
with heteroscedastic volatility
in the
South African financial markets**

MG Ferreira

Presented in partial fulfilment of the degree MSc
in the Faculty of Science
University of Pretoria
October 1999

SOLI DEO GLORIA

Contents

Preface	iii
List of Figures	ix
List of Tables	xi
Chapter 1. Time Series Models	1
1. Introduction	1
2. Stationary Time Series	3
3. Autocovariance and Autocorrelation	15
4. Partial Autocorrelation	19
5. Linear Cyclical Models	22
6. Difference Equations	35
7. General Linear Models	52
8. ARMA Models	55
9. ARIMA Models	98
Chapter 2. Heteroscedastic Time Series Models	103
1. Introduction	103
2. ARCH Models	106
3. GARCH Models	114
Chapter 3. The South African News Impact Curve	129
1. Introduction	129
2. Fitting ARMA Models	131
3. Fitting GARCH Models	137
4. Conclusion	156
Bibliography	159
Subject and Notation Index	163

List of Figures

1.1	JSE Overall Index - Daily Close	12
1.2	JSE Overall Index - Daily Change	13
1.3	JSE Overall Index - Daily Percentage Change	14
1.4	White Noise	23
1.5	The Function $\cos(\theta)$ on the Basic Interval	23
1.6	SAB Quarterly Sales (Millions of Liters)	29
1.7	Percentage Returns (Month on Year)	34
1.8	Random Walk with No Drift	59
1.9	Random Walk with Drift	59
1.10	Parameter Range of Stationary AR(2) Model	73
1.11	Contours of ARMA(1, 1) Variance	94
1.12	Contours of Slope α_1 in $\kappa = \alpha_0 + \alpha_1 \kappa_\varepsilon$	100
1.13	Example Spread and Spread Returns	101
2.1	ZAR/USR Exchange Rate	106
2.2	The Normal and ARCH(1) distributions	112
2.3	GARCH(1, 1) Kurtosis	121
2.4	Arbitrary News Impact Curve for Various Models	127
3.1	Actual South African Market Indices	130
3.2	Transformed Market Indices	130
3.3	Sample Autocorrelation Function	132
3.4	Sample Partial Autocorrelation Function	132
3.5	Ljung-Box Statistic	136
3.6	Bayesian Information Criterion (<i>BIC</i>)	136

3.7	SAC of Squared Errors	139
3.8	SPAC of Squared Errors	139
3.9	MM GARCH Models at Order 2	142
3.10	MM AGARCH News Impact Curve	144
3.11	MM GARCH Models at Order 5	145
3.12	MM GJR News Impact Curve	147
3.13	BM GARCH Models at Order 3	148
3.14	BM GJR News Impact Curve	150
3.15	SM GARCH Models at Order 2	151
3.16	SM GJR News Impact Curve	153
3.17	FM GARCH Models at Order 4	153
3.18	FM GJR News Impact Curve	154

List of Tables

1.1	General Statistics for Time Series	1
1.2	General Definitions for Stationary Time Series	4
1.3	Examples of Transformation Functions	10
1.4	Periodogram of SAB Sales (First Differences)	30
1.5	JSE Overall Index Percentage Returns (Month on Year)	34
1.6	ZAR/USD Percentage Returns (Month on Year)	35
1.7	Sample Coefficients of y , $\bar{y}$ and d with $m = 4$	63
1.8	Inflation Rate in Avalon	77
1.9	Behaviour of ARMA Models used for Model Identification	87
1.10	Expected Terms of Fourth Moment of ARMA(1, 1) Model	97
1.11	Terms in Numerator and Denominator of ARMA(1, 1)	
	Kurtosis	99
2.1	Appropriate Variance Stabilizing Transformations When $\sigma \propto f(\mu)$	105
3.1	Descriptive Statistics for MM, BM, SM and FM	131
3.2	ARMA Model Parameters	138
3.3	MM AGARCH Model Parameters	143
3.4	MM GJR Model Parameters	146
3.5	BM GJR Model Parameters	149
3.6	SM GJR Model Parameters	152
3.7	FM GJR Model Parameters	155

CHAPTER 1

Time Series Models

1. Introduction

Most of statistics deal with stochastic variables that may be fairly accurately simulated using random sampling procedures. This replication process assumes almost always that the observations are independent of one another. In *time series analysis* we can only observe one realization at each point in time and these observations are usually dependent over time.

Thus we implicitly assume that the n observations

$$y_t, \quad t = 0, 1, 2, \dots, n-1$$

in a so called *time series* are dependent on each other and that the nature of these dependencies is of interest in itself. The body of techniques available for the analysis of such a series of dependent observations and the study of these dependencies is called *time series analysis*.

In many cases and for many purposes we only – or at least – require knowledge of the mean, the variance and the autocorrelation structure of the series. For a time series y_t we have the following descriptive statistics:

mean:	$E[y_t] = \mu_t$
variance:	$\text{Var}(y_t) = E[(y_t - \mu_t)^2] = \sigma_t^2$
standard deviation:	$\text{Std}(y_t) = \sqrt{\text{Var}(y_t)} = \sigma_t$
autocovariance:	$\text{Cov}(y_t, y_{t-k}) = E[(y_t - \mu_t)(y_{t-k} - \mu_{t-k})] = \gamma_{t,k}$
autocorrelation:	$\text{Cor}(y_t, y_{t-k}) = \frac{\gamma_{t,k}}{\sigma_t \sigma_{t-k}} = \rho_{t,k}$

TABLE 1.1. General Statistics for Time Series

From Table 1.1 it follows that $\sigma_t^2 = \gamma_{t,0}$, $\sigma_{t-k}^2 = \gamma_{t-k,0}$ and $\rho_{t,0} = 1$. Later on we will also make use of the following property of autocorrelations.

THEOREM 1.1.

$$|\text{Cor}(y_t, y_{t-k})| = |\rho_{t,k}| \leq 1. \quad (1.1)$$

PROOF. To show the result we form the new variable

$$x_t = \frac{y_t}{\sigma_t} - \rho_{t,k} \frac{y_{t-k}}{\sigma_{t-k}} \quad (1.2)$$

with variance

$$\begin{aligned} \text{Var}(x_t) &= \text{E}[x_t - \text{E}[x_t]]^2 \\ &= \text{E}\left[\frac{y_t - \mu_t}{\sigma_t} - \rho_{t,k} \left(\frac{y_{t-k} - \mu_{t-k}}{\sigma_{t-k}}\right)\right]^2 \\ &= \text{E}\left[\frac{y_t - \mu_t}{\sigma_t}\right]^2 + \rho_{t,k}^2 \text{E}\left[\frac{y_{t-k} - \mu_{t-k}}{\sigma_{t-k}}\right]^2 \\ &\quad - 2\rho_{t,k} \text{E}\left[\frac{y_t - \mu_t}{\sigma_t} \cdot \frac{y_{t-k} - \mu_{t-k}}{\sigma_{t-k}}\right] \\ &= 1 + \rho_{t,k}^2 - 2\rho_{t,k}^2 \\ &= 1 - \rho_{t,k}^2 \geq 0. \end{aligned}$$

□

COROLLARY 1.2. *It follows from the definition of autocorrelation that*

$$|\gamma_{t,0}| = \gamma_{t,0} \geq |\gamma_{t,k}|. \quad (1.3)$$

Note that the result also holds if we replace y_t and y_{t-k} with any stochastic variables. When calculating the covariance between two variables it is often more convenient to use

$$\begin{aligned} \text{Cov}(y_t, y_{t-k}) &= \text{E}[y_t y_{t-k}] - 2 \text{E}[y_t] \text{E}[y_{t-k}] + \text{E}[y_t] \text{E}[y_{t-k}] \\ &= \text{E}[y_t y_{t-k}] - \text{E}[y_t] \text{E}[y_{t-k}]. \end{aligned}$$

Finally note that if y_t and y_{t-k} are independent so that

$$\text{E}[y_t y_{t-k}] = \text{E}[y_t] \text{E}[y_{t-k}]$$

then the autocovariance $\gamma_{t,k}$ (and autocorrelation) is zero.

2. Stationary Time Series

In order to develop inferential procedures we must recreate some notion of replicability. We achieve this by the assumption of *stationarity*, which intuitively means that the series fluctuate around a constant mean and variance and that the structure of the series depends only on the relative position in time of the observations.

DEFINITION 1.3. A time series y_t is said to be *strictly stationary* if the joint probability distribution of $y_{t_1}, y_{t_2}, \dots, y_{t_n}$ is the same as the joint probability distribution of $y_{t_1+\tau}, y_{t_2+\tau}, \dots, y_{t_n+\tau}$ for all $t_1, t_2, \dots, t_n, \tau$.

DEFINITION 1.4. A time series y_t is *weakly, second order or covariance stationary*, or simply, *stationary*, if:

1. *Constant Mean:*

If y_t has a constant mean, $E[y_t] = \mu, \quad \forall t$.

2. *Constant Variance:*

If y_t has constant variance, $\text{Var}(y_t) = \sigma^2, \quad \forall t$.

3. *Constant Covariance:*

If the covariance between equally spaced observations is constant and independent of time so that we can write it as

$$\text{Cov}(y_t, y_{t-k}) = \gamma_k, \quad \forall t, k.$$

THEOREM 1.5. *If y_t is covariance stationary then*

$$\gamma_k = \gamma_{-k}. \quad (1.4)$$

PROOF.

$$\begin{aligned} \gamma_k &= E[(y_t - \mu)(y_{t-k} - \mu)] \\ &= E[(y_{t+k} - \mu)(y_{(t-k)+k} - \mu)] \\ &= E[(y_t - \mu)(y_{t+k} - \mu)] \\ &= \gamma_{-k}. \end{aligned}$$

□

From now on, when we refer to stationarity it will imply covariance stationarity. If a time series y_t is stationary we may adjust Table 1.1 to:

mean:	$E[y_t] = \mu$
variance:	$\text{Var}(y_t) = E[(y_t - \mu)(y_t - \mu)] = \sigma^2$
standard deviation:	$\text{Std}(y_t) = \sqrt{\text{Var}(y_t)} = \sigma$
autocovariance:	$\text{Cov}(y_t, y_{t-k}) = E[(y_t - \mu)(y_{t-k} - \mu)] = \gamma_k = \gamma_{-k}$
autocorrelation:	$\text{Cor}(y_t, y_{t-k}) = \frac{\gamma_k}{\sqrt{\sigma^2 \sigma^2}} = \rho_k = \frac{\gamma_k}{\gamma_0}$

TABLE 1.2. General Definitions for Stationary Time Series

2.1. Linear Combinations of Stationary Time Series. Here we investigate the properties of a time series L_t formed as a linear combination of a stationary time series y_t

$$L_t = l_0 y_t + l_1 y_{t-1} + \cdots + l_{n-1} y_{t-(n-1)}. \quad (1.5)$$

THEOREM 1.6. *The mean of a finite linear combination (1.5) of a stationary time series y_t is*

$$E[L_t] = \mu \sum_{i=0}^{n-1} l_i. \quad (1.6)$$

PROOF.

$$\begin{aligned} E[L_t] &= E \left[\sum_{i=0}^{n-1} l_i y_{t-i} \right] \\ &= \sum_{i=0}^{n-1} l_i E[y_{t-i}] \\ &= \mu \sum_{i=0}^{n-1} l_i. \end{aligned}$$

□

THEOREM 1.7. *The covariance of a finite linear combination (1.5) of a stationary time series y_t is*

$$\text{Cov}(L_t, L_{t-k}) = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} l_i l_j \gamma_{k-i+j}. \quad (1.7)$$

PROOF.

$$\begin{aligned} \text{Cov}(L_t, L_{t-k}) &= \text{E} \left[\left(\sum_{i=0}^{n-1} l_i y_{t-i} \right) \cdot \left(\sum_{i=0}^{n-1} l_i y_{t-k-i} \right) \right] \\ &\quad - \text{E} \left[\sum_{i=0}^{n-1} l_i y_{t-i} \right] \cdot \text{E} \left[\sum_{i=0}^{n-1} l_i y_{t-k-i} \right] \\ &= \text{E} [l_0^2 y_t y_{t-k} + l_0 l_1 y_t y_{t-k-1} + \dots \\ &\quad + l_{n-1} l_{n-2} y_{t-n-1} y_{t-k-n-2} + l_{n-1}^2 y_{t-n-1} y_{t-k-n-1}] \\ &\quad - l_0^2 \text{E}[y_t] \text{E}[y_{t-k}] - l_0 l_1 \text{E}[y_t] \text{E}[y_{t-k-1}] - \dots \\ &\quad - l_{n-1} l_{n-2} \text{E}[y_{t-n-1}] \text{E}[y_{t-k-n-2}] \\ &\quad - l_{n-1}^2 \text{E}[y_{t-n-1}] \text{E}[y_{t-k-n-1}] \\ &= \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} l_i l_j [\text{E}[y_{t-i} y_{t-k-j}] - \text{E}[y_{t-i}] \text{E}[y_{t-k-j}]] \\ &= \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} l_i l_j \gamma_{k-i+j}. \end{aligned}$$

□

COROLLARY 1.8. *Since the mean and the covariance are constants independent of t it follows that any finite linear combination of a stationary time series again is stationary.*

COROLLARY 1.9. *In particular, first differences $L_t = y_t - y_{t-1}$ of a stationary time series y_t is stationary.*

COROLLARY 1.10. *Recursively any finite order of differences of a stationary time series is stationary.*

COROLLARY 1.11. *A few examples of (1.7). For $n = 1, 2$ and 3 we have*

$$n = 1 \quad \text{Cov}(L_t, L_{t-k}) = l_0^2 \gamma_k \quad (1.8)$$

$$n = 2 \quad \text{Cov}(L_t, L_{t-k}) = l_0^2 \gamma_k + l_0 l_1 \gamma_{k+1} + l_1 l_0 \gamma_{k-1} + l_1^2 \gamma_k \quad (1.9)$$

$$\begin{aligned} n = 3 \quad \text{Cov}(L_t, L_{t-k}) = & l_0^2 \gamma_k + l_0 l_1 \gamma_{k+1} + l_0 l_2 \gamma_{k+2} \\ & + l_1 l_0 \gamma_{k-1} + l_1^2 \gamma_k + l_1 l_2 \gamma_{k+1} \\ & + l_2 l_0 \gamma_{k-2} + l_2 l_1 \gamma_{k-1} + l_2^2 \gamma_k \end{aligned} \quad (1.10)$$

Although we could have written (1.5) as the more general

$$L'_t = \alpha + l_0 y_t + l_1 y_{t-1} + \cdots + l_{n-1} y_{t-(n-1)} = \alpha + L_t$$

which also includes a constant α we would not have gained from it. Such an *origin shift* only affects the mean

$$E[L'_t] = E[\alpha + L_t] = \alpha + E[L_t]. \quad (1.11)$$

and not the covariance.

2.2. White Noise. One of the best and most simple examples of a stationary time series is the random or white noise series ε_t where the ε_t 's are *identically and independently distributed* or *i.i.d.* with mean zero and constant variance:

$$E[\varepsilon_t] = 0, \quad \forall t$$

$$\text{Var}(\varepsilon_t) = E[\varepsilon_t^2] = \sigma_\varepsilon^2, \quad \forall t$$

$$\text{Cov}(\varepsilon_t, \varepsilon_{t-k}) = \text{Cor}(\varepsilon_t, \varepsilon_{t-k}) = 0, \quad \forall t, k \neq 0.$$

We will from now on call such a zero mean i.i.d. series *white noise*¹. We can easily generate white noise through taking repeated samples from some distribution and we also say that the white noise *process* is *generated* by an e.g. normal distribution.

¹White light is made up of several light waves of random frequencies. Similarly noise consists of sound waves of various random frequencies and amplitudes.

We already know that a finite linear combination of stationary white noise will again be stationary but will now investigate the stationarity conditions of an *infinite* combination. Since we have for i.i.d. white noise

$$\gamma_k = \begin{cases} \sigma_\varepsilon^2 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

we start by simplifying (1.7).

COROLLARY 1.12. *A linear combination L_t of white noise ε_t has*

$$\begin{aligned} \text{Cov}(L_t, L_{t-k}) &= \sum_{\substack{0 \leq i, j \leq n-1 \\ j-i=k}} l_i l_j \gamma_0 \\ &= \begin{cases} \gamma_0 \cdot \sum_{i=0}^{n-1-k} l_i l_{i+k} & |k| \leq n-1 \\ 0 & |k| > n-1 \end{cases}. \end{aligned} \quad (1.12)$$

This leads us to the main result.

THEOREM 1.13. *An infinite linear combination of white noise is stationary if, and only if, the weights l_i are square summable*

$$\sum_{i=0}^{\infty} l_i^2 < \infty.$$

PROOF. The mean of such a linear combination from (1.6) is

$$\begin{aligned} E[L_t] &= \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} l_i \mu_\varepsilon \\ &= 0. \end{aligned}$$

The autocovariance from (1.12) becomes

$$\text{Cov}(L_t, L_{t-k}) = \gamma_0 \cdot \lim_{n \rightarrow \infty} \left(\sum_{i=0}^{n-1-k} l_i l_{i+k} \right).$$

Corollary 1.2 requires us to only show that $\text{Cov}(L_t, L_t)$ is finite. Since

$$\text{Cov}(L_t, L_t) = \sigma_\varepsilon^2 \cdot \lim_{n \rightarrow \infty} \left(\sum_{i=0}^{n-1} l_i^2 \right) \quad (1.13)$$

the result follows immediately. $\square$

COROLLARY 1.14. *An infinite linear combination of white noise is stationary if the weights are absolute summable*

$$\sum_{i=0}^{\infty} |l_i| < \infty.$$

PROOF. We only need to show that *absolute* is a stronger requirement than *square summability*. If l_i is absolute summable then

$$\lim_{i \rightarrow \infty} l_i = 0$$

and there exists a finite N so that $|l_i| < 1$, $\forall i \geq N$ which implies that $l_i^2 < |l_i|$. Now

$$\begin{aligned} \sum_{i=0}^{\infty} l_i^2 &= \sum_{i=0}^{N-1} l_i^2 + \sum_{i=N}^{\infty} l_i^2 \\ &< \sum_{i=0}^{N-1} l_i^2 + \sum_{i=N}^{\infty} |l_i| < \infty. \end{aligned}$$

□

The converse is not true – a square summable series is not necessarily absolutely summable. The prime example is

$$l_i = \frac{1}{i}$$

which, as is well known (e.g. Ellis & Gulick, 1986:514), is square but not absolutely summable.

COROLLARY 1.15. *An infinite linear combination of white noise with weights*

$$l_i = \psi^i$$

is stationary if, and only if, $|\psi| < 1$.

PROOF. The series ψ^i is summable only if $|\psi| < 1$. Since $l_i \leq |l_i|$ this implies absolute summability, with sum, as is well known (e.g. Ellis & Gulick, 1986:542)

$$\sum_{i=0}^{\infty} l_i = \sum_{i=0}^{\infty} \psi^i = \frac{1}{1 - \psi}.$$

□

Note that white noise is also strictly stationary. Since ε_t is i.i.d. we can write the joint distribution function as

$$\begin{aligned} f_{\varepsilon_{t_1}, \varepsilon_{t_2}, \dots, \varepsilon_{t_n}}(\varepsilon_{t_1}, \varepsilon_{t_2}, \dots, \varepsilon_{t_n}) &= f_{\varepsilon_{t_1}}(\varepsilon_{t_1}) \cdot f_{\varepsilon_{t_2}}(\varepsilon_{t_2}) \cdot \dots \cdot f_{\varepsilon_{t_n}}(\varepsilon_{t_n}) \\ &= f(\varepsilon_{t_1}) \cdot f(\varepsilon_{t_2}) \cdot \dots \cdot f(\varepsilon_{t_n}) \\ &\propto f(\varepsilon_{t_1+\tau}) \cdot f(\varepsilon_{t_2+\tau}) \cdot \dots \cdot f(\varepsilon_{t_n+\tau}). \end{aligned}$$

2.3. Transforming Non Stationary to Stationary Time Series.

Although many time series – such as stock prices – actually encountered in practice exhibit *non stationary* behaviour such series often exhibit homogeneous behaviour of some kind. In order to *extract* this homogeneous behaviour we make use of some transformation or series of transformations

$$\tilde{y}_t = T(y_t) = T_1 \circ T_2 \circ \dots \circ T_n(y_t)$$

in order to *detrend* the series and often also to remove increasing variation over time.

2.3.1. *Differencing and the Backward Shift Operator.* We have already seen that differencing a stationary time series yields another stationary time series. Box and Jenkins (1976:7) recommended applying differencing in order to detrend a non stationary time series. The easiest way to formulate the process of differencing is by making use of the *backward shift operator* B defined as

$$B y_t = y_{t-1}. \quad (1.14)$$

This operator may be applied repeatedly in order to difference a time series to the order d as follows:

$$\begin{aligned} d = 1 & : & (1 - B)y_t &= y_t - y_{t-1} \\ d = 2 & : & (1 - B)^2 y_t &= y_t - 2y_{t-1} + y_{t-2} \\ d = 3 & : & (1 - B)^3 y_t &= y_t - 3y_{t-1} + 3y_{t-2} - y_{t-3} \\ & \vdots & & \end{aligned}$$

T	Transformation
$\tilde{y}_t = (1 - B)^d y_t$	Differencing
$\tilde{y}_t = y_t - \hat{y}_t$	Regression model ($\hat{y}_t = \beta \mathbf{X}_t$)
$\tilde{y}_t = \frac{y_t}{y_{t-1}} - 1$	Percentage changes or returns
$\tilde{y}_t = \ln \frac{y_t}{y_{t-1}}$	Log returns

TABLE 1.3. Examples of Transformation Functions

A non stationary time series that becomes stationary after application of the backward shift operator is referred to as a *difference stationary* time series. Other often used transformation functions are provided in Table 1.3.

The regression model in Table 1.3 will usually be some polynomial in time but may also be a deterministic trend function such as an econometric model. If a time series is stationary around a deterministic time trend then it is described as a *trend stationary* time series. A stationary time series is obtained upon subtracting the trend.

There are significant differences in the forecasts, confidence intervals and dynamic multipliers that are obtained from a trend stationary and difference stationary time series. A trend stationary forecast converges to the trend itself and with a finite confidence interval. On the other hand a difference stationary forecast converges to the trend *projected* from the last observed value of the time series while the confidence interval eventually grows linearly with time (Hamilton 1994:435-444).

If a trend stationary time series is differenced the result is a stationary but non invertible² series. On the other hand subtracting a time trend from a difference stationary time series does remove the trend but the variance will still increase with time (Hamilton 1994:444). There have been a considerable debate in the econometric literature as to whether trends should be removed from non stationary time series using a polynomial in time regression model

²We will discuss the concept of invertibility later.

or through the use of differencing (Nelson & Kang 1980). At this stage the evidence strongly supports the use of differencing rather than polynomial regression models.

2.3.2. Calculating Returns. When working with financial time series, especially if they record prices of financial instruments, the most often utilized transformation is that of calculating either percentage or log returns. Mathematicians normally make use of log returns due to the bulk of known mathematical properties available while practitioners and market commentators often use percentage returns.

It is easy to see why these transformations have such great importance. The daily returns of an investor who invested a given amount in a financial instrument is reflected by the percentage return applied to the original investment. Having bought 100 shares at a price of R100 (initial investment of R10,000) will yield a profit of 1% if the price increases with 1% since the value of the investment grows to $R101 \times 100$ or with 1% to R10,100.

The same percentage increase (or decrease) is felt in the investment as in the price itself and, since a percentage can easily be applied to any amount, it is very comfortable to express price movements as percentages. Since asset prices do not assume negative values it is also not necessary to cater for the negative range where the percentage returns and log returns transformations do not make sense.

There is not a big difference between the log returns and the percentage returns transformations. From the well known property (e.g. Ellis & Gulick, 1986:568)

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots, \quad \forall |x| < 1$$

it follows that

$$\ln\left(\frac{y_t}{y_{t-1}}\right) = \ln\left[1 + \left(\frac{y_t}{y_{t-1}} - 1\right)\right] \approx \frac{y_t}{y_{t-1}} - 1$$

especially if $\frac{y_t}{y_{t-1}}$ is small which we will expect if the duration Δt is small. Since daily returns are usually close to zero and much less than 100% these two approaches yield very similar results in practice.

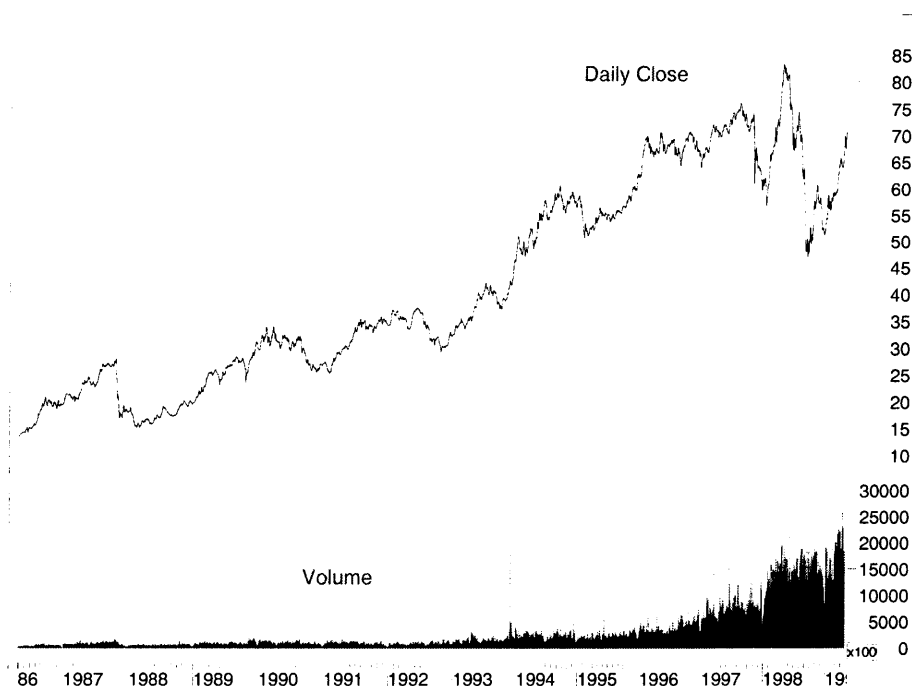


FIGURE 1.1. JSE Overall Index - Daily Close

2.3.3. *Seasonality.* If there is also *seasonality* present in the time series, such as will be the case with many commodity prices, it often helps to adjust the time span inherent in the transformation in order to accommodate this seasonality. Thus the transformation functions given in Table 1.3 for the percentage changes for example will change to $\frac{y_t}{y_{t-d}} - 1$ where d is a suitable value reflecting the seasonality believed to be present in the series.

A typical example is *consumer inflation* figures where the percentage change in the actual series (called the *consumer price index* or *CPI*) is calculated on a twelve month rolling basis. Instead of calculating the percentage change in the index from the previous month, the percentage change from the same month a year before is calculated ($d = 12$). This so called *year on year* percentage change is reported in the press as the *headline inflation*. It has the added advantage of being an *annual* figure which may readily be compared to other similar figures such as annual salary increases (usually lower) or interest rates (usually higher).

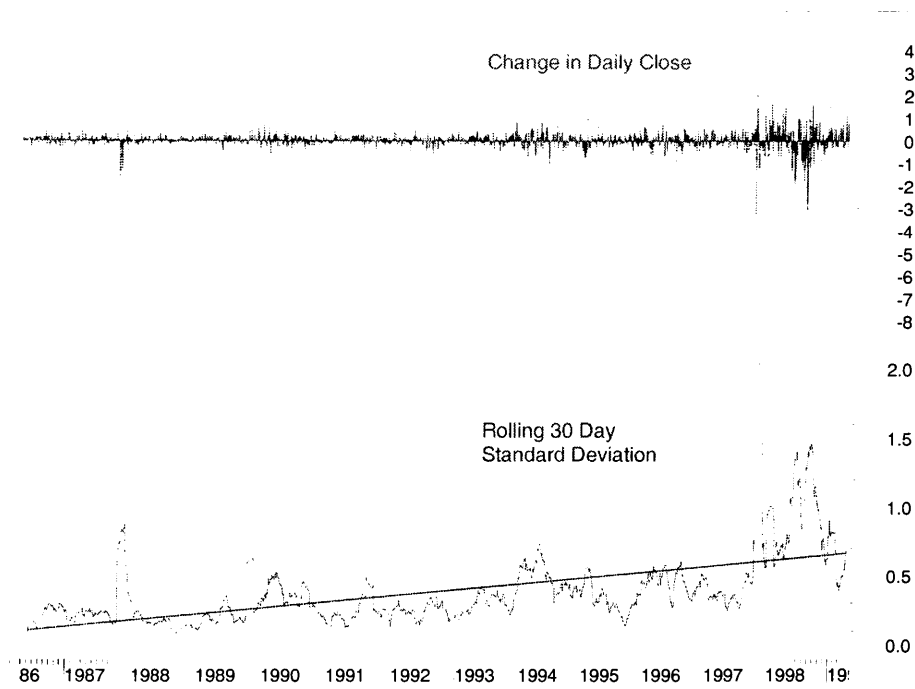


FIGURE 1.2. JSE Overall Index - Daily Change

2.3.4. *A Typical Financial Time Series.* Notice in Figure 1.1 how the trajectory of the time series exhibits a clear long term upward trend. This time series is not stationary and needs to be detrended using a suitable function from Table 1.3. After taking first differences the time series in Figure 1.2 is obtained. The new time series appears to have a constant mean and is a lot more stationary than before. Note however that the variance of the new series does not remain constant as time advances. The time series appears to *funnel out* or increase in magnitude in the latter part of Figure 1.2. This becomes especially evident if some measure of variance such as a rolling standard deviation is calculated as in the figure. The linear regression fitted to this standard deviation also amplifies the funneling out phenomena.

We can also calculate the percentage returns in stead of the daily changes. This transformation is usually required for a time series exhibiting a long term exponential growth trend – a very common occurrence in financial time

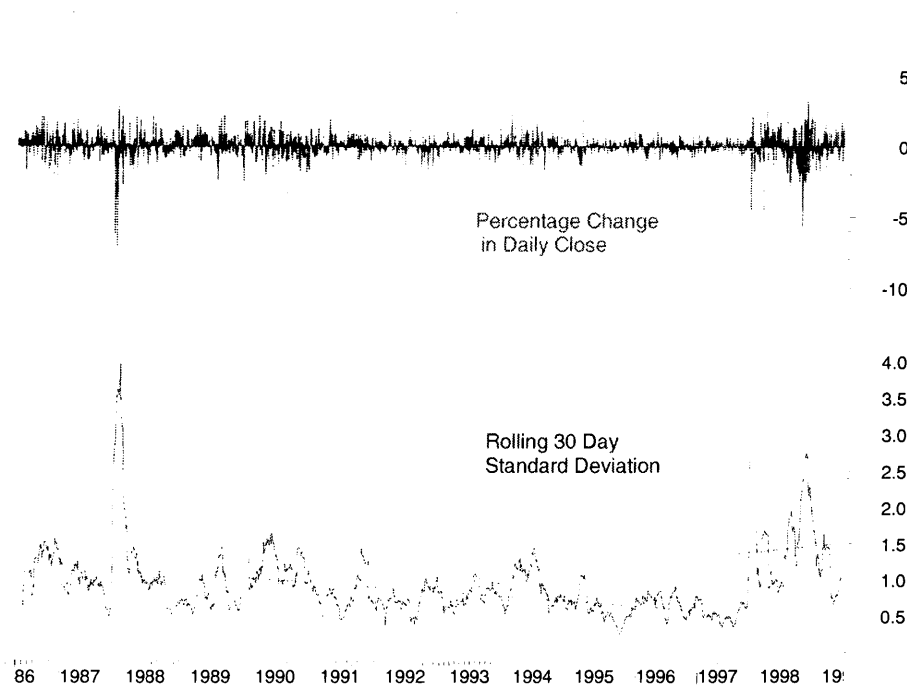


FIGURE 1.3. JSE Overall Index - Daily Percentage Change

series that record the price of some asset. This transformation then removes both the trend and the increasing variance. A quick inspection of Figure 1.1 naturally suggests this approach.

The daily percentage returns of the JSE Overall Index is depicted in Figure 1.3. This series again appears to be a lot more stationary than before, but if it is inspected closer it does not seem like such a plausible assumption either. Again the mean appears to be fairly constant but the variance *clusters* together in certain areas as also pointed out by the various regression lines in the figure – the last few years was for example much more turbulent than previous years.

Even though these remarks are rather general and requires closer investigation they are also quite typical of financial time series.

3. Autocovariance and Autocorrelation

It is well known that a significant positive covariance between two variables indicates a significant positive linear relationship and vice versa. Insignificant covariance similarly indicates the lack of a linear relationship. Since it is not easy to know if such a number is significant we normally calculate the correlation which as we have seen is always less than 1 in magnitude. One logical interpretation of correlation is that it measures the *strength* of the linear relationship between two variables. Notice from (1.2) that if $\rho_{t,k} = \pm 1$ the variance of x_t is zero and y_t is simply a linear function of y_{t-k} .

3.1. The Autocovariance and Autocorrelation Matrices. When working with time series we often need to study a large number of autocovariances and autocorrelations in order to investigate the structure of the series. Then it is helpful to contain these statistics in matrices.

DEFINITION 1.16. The (symmetric) autocovariance $\mathbf{\Gamma}_k$ and autocorrelation $\mathbf{P}_k$ matrices associated with a stationary time series y_t is defined as:

$$\begin{aligned}\mathbf{\Gamma}_k &= \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{k-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{k-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{k-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \gamma_{k-1} & \gamma_{k-2} & \gamma_{k-3} & \cdots & \gamma_0 \end{pmatrix} \\ &= \sigma^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{pmatrix} = \sigma^2 \mathbf{P}_k.\end{aligned}$$

We do not normally calculate all possible covariances and correlations as they usually become insignificant after a few lags as we shall see later. If a time series has insignificant autocovariance of order k and higher we simply reduce the size of the autocovariance and autocorrelation matrices accordingly.

DEFINITION 1.17. *Positive Definiteness of Symmetric Matrices*

Let $\mathbf{A}$ be a real symmetric matrix of order n . Then $\mathbf{A}$ is said to be *positive definite* if

$$\mathbf{x}'\mathbf{A}\mathbf{x} > 0, \quad \forall \mathbf{x} \in \mathbb{R}^n, \mathbf{x} \neq \mathbf{0}.$$

We say $\mathbf{A}$ is *positive semidefinite* if

$$\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0, \quad \forall \mathbf{x} \in \mathbb{R}^n, \mathbf{x} \neq \mathbf{0}$$

where $\mathbf{x}'$ is the transpose of $\mathbf{x}$ in both cases.

Positive definiteness of a matrix gives rise to various useful results such as positive roots, non singularity and so forth. We only give the ones we need here with short proofs.

THEOREM 1.18. *If a matrix is positive definite then all the diagonal elements $a_{ii} > 0$ while if it is positive semi-definite all the diagonal elements $a_{ii} \geq 0$.*

PROOF. With $\mathbf{x}_i = [0, 0, \dots, 0, 1, 0, \dots, 0]$ we have $a_{ii} = \mathbf{x}_i'\mathbf{A}\mathbf{x}_i > 0$ if $\mathbf{A}$ is positive definite and similarly $a_{ii} \geq 0$ for $\mathbf{A}$ positive semidefinite. $\square$

THEOREM 1.19. *If a matrix $\mathbf{A}$ is positive definite there exists a triangular matrix $\mathbf{T}$ with transpose $\mathbf{T}'$ such that*

$$\mathbf{A} = \mathbf{T}'\mathbf{T} \tag{1.15}$$

PROOF. Suppose that $\mathbf{T}'$ is a lower triangular matrix

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} t_{11} & 0 & \cdots & 0 \\ t_{21} & t_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ t_{n1} & t_{n2} & \cdots & t_{nn} \end{pmatrix} \begin{pmatrix} t_{11} & t_{21} & \cdots & t_{n1} \\ 0 & t_{22} & \cdots & t_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & t_{nn} \end{pmatrix}.$$

For $\mathbf{T}$ to exist we must be able to solve these implied equations

$$\begin{aligned} t_{11}^2 &= a_{11}, t_{11}t_{21} = a_{12}, t_{11}t_{31} = a_{13}, \dots, t_{11}t_{n1} = a_{1n} \\ t_{21}t_{11} &= a_{21}, t_{21}^2 + t_{22}^2 = a_{22}, t_{21}t_{31} + t_{22}t_{32} = a_{23}, \dots, t_{21}t_{n1} + t_{22}t_{n2} = a_{2n} \\ &\vdots \\ t_{n1}t_{11} &= a_{n1}, t_{n1}t_{21} = a_{n2}, \dots, \sum_{i=1}^n t_{ni}^2 = a_{nn}. \end{aligned}$$

Since $a_{ii} > 0$ we can solve this recursive set of equations as follows:

$$\begin{aligned} t_{11} &= \pm\sqrt{a_{11}}, t_{i1} = \frac{a_{i1}}{t_{11}}, \dots, \\ t_{21} &= \frac{a_{21}}{t_{11}}, t_{22} = \pm\sqrt{a_{22} - t_{21}^2}, t_{i2} = \frac{a_{21} - t_{21}t_{i1}}{t_{22}} \\ &\vdots \end{aligned}$$

This shows that we can obtain $\mathbf{T}$ in terms of the elements of $\mathbf{A}$. $\square$

COROLLARY 1.20. *The determinant and all principle minors of a positive definite matrix is greater than zero.*

PROOF. Writing the determinant as $|\mathbf{A}| = |\mathbf{T}'| \cdot |\mathbf{T}|$ yields the result. $\square$

THEOREM 1.21. *The autocovariance matrix of a stationary time series is positive definite.*

PROOF. The variance of a linear combination of the stationary y_t

$$\text{Var}(L_t) = \text{Var}(l_0 y_t + l_1 y_{t-1} + \dots + l_{n-1} y_{t-(n-1)})$$

will be greater than zero if not all the l_i 's are zero. Since y_t is stationary with $\gamma_k = \gamma_{-k}$ we have from (1.7) with $k = 0$

$$\begin{aligned} \text{Var}(L_t) &= \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} l_i l_j \gamma_{|j-i|} \\ &= \mathbf{l}' \mathbf{\Gamma}_k \mathbf{l}. \end{aligned}$$

$\square$

COROLLARY 1.22. *The autocorrelation matrix $\mathbf{P}_k = \frac{1}{\sigma^2} \mathbf{\Gamma}_k$ is positive definite.*

We can therefore apply the known properties of positive definite matrices to the autocovariance and autocorrelation matrices. We have for example from Corollary 1.20 and for $n = 2$

$$\begin{aligned} \begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix} &> 0 \\ 1 - \rho_1^2 &> 0 \\ |\rho_1| &< 1. \end{aligned} \tag{1.16}$$

Similarly for $n = 3$ we have:

$$\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix} > 0, \quad \begin{vmatrix} 1 & \rho_2 \\ \rho_2 & 1 \end{vmatrix} > 0, \quad \begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix} > 0$$

yielding the relations

$$|\rho_1| < 1, \quad |\rho_2| < 1, \quad 2\rho_1^2 - \rho_2 < 1 \tag{1.17}$$

and so on.

3.2. The Autocovariance Generating Function. A convenient way of obtaining the autocovariances of a linear combination L_t is found in the autocovariance generating function.

DEFINITION 1.23. The autocovariance generating function of the stationary time series y_t is defined as

$$\gamma_y(z) = \sum_{j=-\infty}^{\infty} \gamma_j z^j. \tag{1.18}$$

Note that in (1.18) the variance of the time series is the coefficient of $z^0 = 1$ and the covariance at lag k is the coefficient of both z^j and z^{-j} . If we have for example the infinite linear combination of white noise

$$L_t = l_0 \varepsilon_t + l_1 \varepsilon_{t-1} + l_2 \varepsilon_{t-2} + \dots$$

then, from (1.12)

$$\begin{aligned}
 \gamma_{L_t}(z) &= \gamma_0 \sum_{j=-\infty}^{\infty} \sum_{i=0}^{\infty} l_i l_{i+j} z^j \\
 &= \gamma_0 \sum_{i=0}^{\infty} \sum_{j=-i}^{\infty} l_i l_{i+j} z^j && \text{since } l_{i+j} = 0 \text{ for } i+j < 0 \\
 &= \gamma_0 \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} l_i l_k z^{k-i} && \text{with } k = i+j \\
 &= \gamma_0 \sum_{k=0}^{\infty} l_k z^k \sum_{i=0}^{\infty} l_i z^{-i}.
 \end{aligned} \tag{1.19}$$

3.3. The Autocorrelation Function.

DEFINITION 1.24. The plot or listing of the values ρ_k is known as the *autocorrelation function* or *ACF*.

If we have a sample $y_0, y_1, \dots, y_{n-1}$ of n observations we may calculate the *Sample Autocorrelation Function* or *SAC* at lag k as

$$r_k = \hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=k}^{n-1-k} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=0}^{n-1} (y_t - \bar{y})^2}.$$

The standard error of the estimate r_k is (Bowerman & O'Connell 1987:34)

$$s_{r_k} = \sqrt{\frac{1 + 2 \sum_{j=1}^{k-1} r_j^2}{n}}.$$

We can use s_{r_k} to conduct hypothesis tests involving the significance of r_k .

The t statistic, commonly used in such tests, will be

$$t_{r_k} = \frac{r_k}{s_{r_k}}.$$

4. Partial Autocorrelation

If successive time series values are highly correlated then there will also be a strong correlation present in values spaced two and even more periods apart. We know that all autocorrelations lie within the range $[-1; +1]$ and from (1.17) it also follows that $\rho_2 > 2\rho_1^2 - 1$. If e.g. $\rho_1 = 0.9$ then we must have $\rho_2 > 0.62$. It makes sense to remove this *carry over effect* through

stripping the ACF at lag k of all intermediary autocorrelations and calculate the *partial autocorrelation* at lag k .

If we want to study for example the influence of y_{t-2} on y_t we have to account for the influence of y_{t-2} on y_{t-1} and consequently the influence of y_{t-1} on y_t . If we strip the autocorrelation ρ_2 between y_t and y_{t-2} of this influence we obtain the partial autocorrelation ϕ_{22} defined as

$$\phi_{22} = \text{Cor}(y_t, y_{t-2})|y_{t-1} = \rho_2|\rho_1.$$

Since there are no intervening observations between y_t and y_{t-1} the first order partial autocorrelation is the same as the autocorrelation

$$\phi_{11} = \rho_1.$$

DEFINITION 1.25. In general we define the partial autocorrelation or *PAC* at lag k as

$$\begin{aligned}\phi_{kk} &= \text{Cor}(y_t, y_{t-k})|\{y_{t-1}, y_{t-2}, \dots, y_{t-k+1}\} \\ &= \rho_k|\{y_{t-1}, y_{t-2}, \dots, y_{t-k+1}\}\end{aligned}$$

or even

$$= \rho_k|\{\rho_1, \rho_2, \dots, \rho_{k-1}\}.$$

The k th partial autocorrelation is the last coefficient in a linear projection of y_t on its k most recent values (Hamilton 1994:111) and is obtained from solving the *Yule-Walker* equations

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix}$$

or, in matrix notation

$$\mathbf{P}_k \boldsymbol{\phi}_k = \boldsymbol{\rho}_k \quad (1.20)$$

successively for $k = 1, 2, \dots$.

The first three partial autocorrelations will be:

$$\phi_{11} = \rho_1 \quad (1.21)$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \quad (1.22)$$

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}. \quad (1.23)$$

In general for ϕ_{kk} the determinant in the numerator has the same elements as that in the denominator but with the last column replaced by

$$\boldsymbol{\rho}_k = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{pmatrix}.$$

4.1. The Partial Autocorrelation Function.

DEFINITION 1.26. The *partial autocorrelation function* or *PAC* is defined as the plot or listing of the partial autocorrelations ϕ_{kk} .

If we have n observations of a time series y_t we calculate the *sample partial autocorrelation function* or *SPAC* through the recursive equation

$$r_{kk} = \begin{cases} r_1, & k = 1 \\ \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j}, & k > 1 \end{cases} \quad (1.24)$$

where

$$r_{kj} = r_{k-1,j} - r_{kk} r_{k-1,k-j}.$$

This equation, which also holds for ρ_{kk} , is due to Durbin (1960) and the procedure is generally known as the *Durbin-Levinson* algorithm. The standard error of r_{kk} is approximately (Bowerman & O'Connell 1987:43)

$$s_{r_{kk}} \simeq \frac{1}{\sqrt{n}}$$

which we use to calculate the t-statistic

$$t_{r_{kk}} = \frac{r_{kk}}{s_{r_{kk}}}$$

that is utilized in hypothesis tests involving ϕ_{kk} .

5. Linear Cyclical Models

If stationary white noise is plotted over time it appears jagged and uneven as in Figure 1.4³. Fortunately most actual time series have a more smooth appearance, due (in part) to the presence of autocorrelation. The closer ρ_1 is to +1 the stronger the linear relationship between y_t and y_{t-1} and the more smooth the series will be – until it becomes a straight line at $\rho_1 = +1$ itself. A notable exception that lacks such smoothness is found in detrended high frequency financial time series – such as daily percentage returns.

Historically the first models to be considered to model autocorrelation tried to estimate the interdependencies in the time series through the introduction of periodical cycles.

DEFINITION 1.27. A function $P(t)$ is periodic with period p if

$$P(t + kp) = P(t), \quad \forall \text{ integers } k.$$

Prime examples of periodic functions are the trigonometric sin and cos functions. These functions also play an important part in the theory of periodic models.

³Computer generated random numbers evenly distributed between 0 and 1.

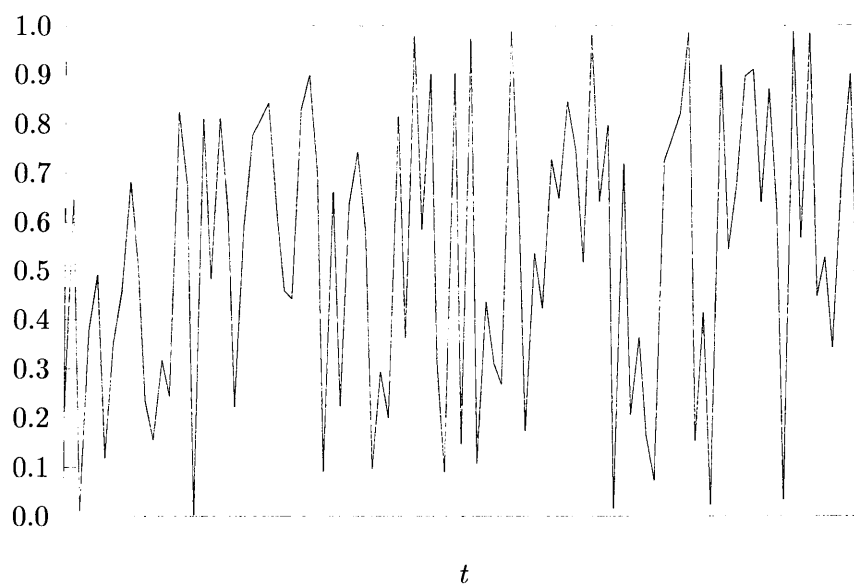
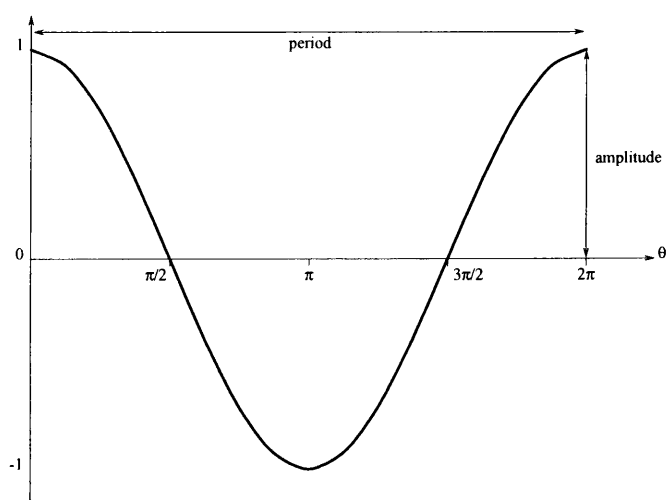


FIGURE 1.4. White Noise

FIGURE 1.5. The Function $\cos(\theta)$ on the Basic Interval

5.1. The Basic Wave. Figure 1.5 displays a plot of the trigonometric function

$$y = \cos(\theta) \quad (1.25)$$

on the *basic interval* $[0, 2\pi]$ with θ in radians.

5.1.1. *Amplitude.* The *amplitude* of (1.25) is 1 but we can easily alter this to any desired number d by simply multiplying with d

$$y = d \cos(\theta).$$

5.1.2. *Period.* One *period*, *wavelength* or *cycle* is the distance that we have to travel along the curve before we start repeating ourselves e.g. from crest to crest. We can alter the period of (1.25) to any desired period p through

$$y = \cos\left(\theta \frac{2\pi}{p}\right).$$

5.1.3. *Frequency.* The *frequency* is the number of full cycles we have on a given interval such as the basic interval. There is one full cycle on the basic interval but we can change this to any amount f through

$$y = \cos(f\theta).$$

The reciprocal relationship between frequency and period on the basic interval is $f = \frac{2\pi}{p}$. If the interval is defined more generally as h (e.g. one week or one quarter) then this relationship becomes $f = \frac{h}{p}$.

5.1.4. *Phase.* The basic wave (1.5) starts with its crest at zero but this can be altered through an origin shift known as the *phase angle*

$$y = \cos(\theta - \alpha).$$

This definition presents a slight problem as we can for example increase the frequency and in so doing decrease the phase angle. To side-step this the phase angle is defined as the distance from the origin to the crest regardless of the frequency and under the assumption that a single wave is completed on the basic interval. The distance from the origin to the crest taking the frequency into consideration is then defined as the *phase shift*.

5.2. The Simple Periodic Model. Armed with these tools we can proceed to the basic building block of the linear cyclical model.

DEFINITION 1.28. The simple periodic model, consisting of one periodic wave, is defined as

$$\begin{aligned} y_t &= P(t) + \varepsilon_t \\ &= d \cos(\omega t - \theta) + \varepsilon_t. \end{aligned} \quad (1.26)$$

We model y_t as a periodic wave on the interval $t = 0, 1, \dots, n - 1$ with amplitude d , frequency ω , phase shift⁴ θ and error ε_t which we assume to be white noise.

LEMMA 1.29. *We can rewrite the simple periodic model as*

$$y_t = a \sin(\omega t) + b \cos(\omega t) + \varepsilon_t \quad (1.27)$$

with

$$a = d \sin(\theta) \quad (1.28)$$

$$b = d \cos(\theta) \quad (1.29)$$

and

$$d^2 = a^2 + b^2. \quad (1.30)$$

PROOF. It follows by the trigonometric identity

$$\cos(A + B) = \cos(A) \cos(B) - \sin(A) \sin(B) \quad (1.31)$$

□

⁴Granger and Newbold (1986:10) as well as Dhrymes (1984:448) defined the simple periodic model as

$$y_t = d \cos(\omega t + \theta) + \varepsilon_t$$

which differs from (1.26) in that it adds rather than subtracts the phase shift θ . This does not affect any of the further results since the phase shift is assumed to be somewhere in the interval $[-\pi, \pi]$. Other texts on either the simple periodic model (Kendall & Ord 1973:156) or Fourier analysis (Rayner 1971:16) subtracted θ and, from the definition of the phase shift, it also makes more sense and this convention have been adopted here.

We will assume that θ in this model is evenly distributed on the interval $[-\pi, \pi]$ so that $E[\theta] = 0$, $E[\sin(\theta)] = 0$ and $E[\cos(\theta)] = 0$. Under this assumption we can prove the following:

THEOREM 1.30. *The mean of the simple periodic model is*

$$\mu_t = 0. \quad (1.32)$$

PROOF.

$$E[y_t] = d \sin(\omega t) E[\sin(\theta)] + d \cos(\omega t) E[\cos(\theta)] + E[\varepsilon_t] = 0.$$

□

THEOREM 1.31. *The autocovariance of the simple periodic model is*

$$\gamma_k = \begin{cases} \frac{d^2}{2} + \sigma_\varepsilon^2, & k = 0 \\ \frac{d^2}{2} \cos(k\omega), & k > 0 \end{cases}. \quad (1.33)$$

PROOF. Using the model (1.26) and the trigonometric identity

$$\cos(A) \cos(B) = \frac{1}{2} (\cos(A + B) + \cos(A - B))$$

we find

$$\begin{aligned} \gamma_k &= E[y_t y_{t-k}] - E[y_t] E[y_{t-k}] \\ &= E\left[(d \cos(\omega t - \theta) + \varepsilon_t)(d \cos(\omega(t-k) - \theta) + \varepsilon_{t-k})\right] - 0 \\ &= E[d^2 \cos(\omega t - \theta) \cos(\omega(t-k) - \theta)] + \sigma_\varepsilon^2 \\ &= \frac{d^2}{2} E[\cos(2\omega t - \omega k - 2\theta) + \cos(\omega k)] + \sigma_\varepsilon^2 \\ &= \begin{cases} \frac{d^2}{2} E[\cos(2\omega t - 2\theta) + 1] + \sigma_\varepsilon^2, & k = 0 \\ \frac{d^2}{2} E[\cos(\omega(2t-k) - 2\theta) + \cos(k\omega)], & k > 0 \end{cases} \\ &= \begin{cases} \frac{d^2}{2} + \sigma_\varepsilon^2, & k = 0 \\ \frac{d^2}{2} \cos(k\omega), & k > 0 \end{cases}. \end{aligned}$$

□

COROLLARY 1.32. *It follows immediately that the simple periodic model is stationary.*

5.3. The Linear Cyclical Model. It is quite natural to suspect a time series to consist of not one but of the sum of many, say m , different cycles $P_i(t)$. Extending (1.26) in this way yields the linear cyclical model.

DEFINITION 1.33. The linear cyclical model of a time series y_t is defined as

$$y_t = \sum_{i=1}^m P_i(t) + \varepsilon_t. \quad (1.34)$$

where P_i is a simple periodic model of period p_i

$$P_i(t) = d_i \cos(\omega_i t - \theta_i) \quad (1.35)$$

and ε_t is white noise.

LEMMA 1.34. *We can rewrite the linear cyclical model as*

$$y_t = \sum_{i=1}^m \left[a_i \sin(\omega_i t) + b_i \cos(\omega_i t) \right] + \varepsilon_t \quad (1.36)$$

with

$$a_i = d_i \sin(\theta_i) \quad (1.37)$$

$$b_i = d_i \cos(\theta_i) \quad (1.38)$$

and

$$d_i^2 = a_i^2 + b_i^2. \quad (1.39)$$

PROOF. This follows from (1.27). $\square$

Previously we assumed that θ is evenly distributed and we maintain this assumption for θ_i . We also assume that θ_i is independent of θ_j for all $i \neq j$.

THEOREM 1.35. *The mean of the linear cyclical model is*

$$\mu_t = 0. \quad (1.40)$$

PROOF. Since the mean of P_k is zero we have

$$E[y_t] = \sum_{i=1}^m E[P_i(t)] + E[\varepsilon_t] = 0.$$

$\square$

THEOREM 1.36. *The autocovariance of the linear cyclical model is*

$$\gamma_k = \begin{cases} \frac{1}{2} \sum_{i=1}^m d_i^2 \cos(k\omega_i) + \sigma_\varepsilon^2, & k = 0 \\ \frac{1}{2} \sum_{i=1}^m d_i^2 \cos(k\omega_i), & k > 0 \end{cases}. \quad (1.41)$$

PROOF. Since the θ_i are independent we have

$$E\left[(\cos(\omega_i t - \theta_i))(\cos(\omega_j t - \theta_j))\right] = 0$$

and the proof is the same as that of Theorem 1.31. $\square$

5.4. Periodogram Analysis. The important parameters in the linear cyclical model (1.34) are the number of periods m and the periods p_k themselves. If this is known it is not difficult to estimate the parameters d_i and θ_i . If these periods are not known there is a number of methods of finding them, known as *periodogram analysis*. The classical and most used method is the Schuster periodogram due to Sir Arthur Schuster (1898:13-17) who popularized it 100 years ago.

DEFINITION 1.37. Given a time series y_t of n observations with n odd, the quantity $I_n(\lambda)$, called the *intensity* at frequency λ , is defined as

$$I_n(\lambda) = \frac{1}{n^2} \left[\left(\sum_{t=0}^{n-1} y_t \cos(t\lambda) \right)^2 + \left(\sum_{t=0}^{n-1} y_t \sin(t\lambda) \right)^2 \right]. \quad (1.42)$$

DEFINITION 1.38. The *Schuster periodogram* is the plot of the intensity against $\frac{2\pi}{\lambda}$.

DEFINITION 1.39. It is often more convenient to plot $I_n(\lambda)$ against λ and this is called the *spectrogram* of the time series.

Both these plots should have clear peaks at the corresponding periods p_k of the linear cyclical model for reasons that will become apparent later on.

Figure 1.6 depicts a financial time series exhibiting obvious cyclical behaviour. The corresponding spectrogram for this series, after taking first differences to remove the trend, is given in Table 1.4. Note from the pulse diagram in the last column of the table the peak at around four quarters.

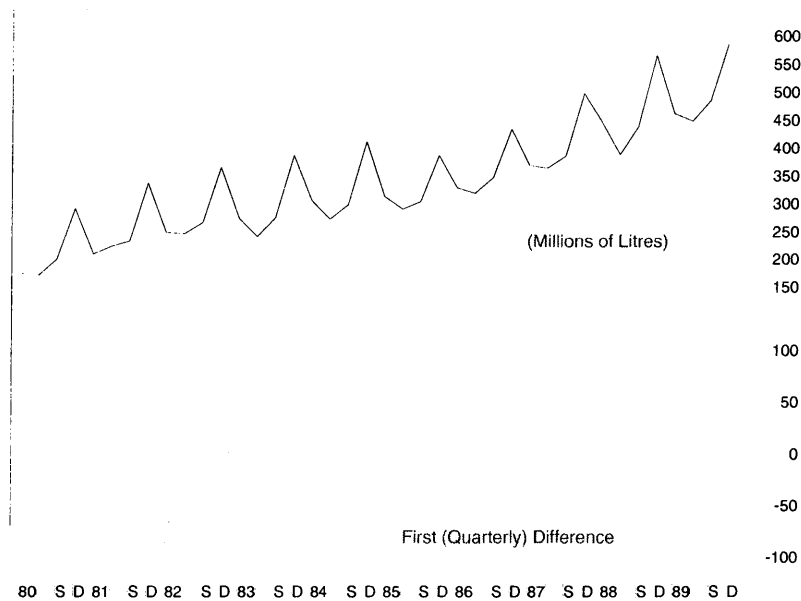


FIGURE 1.6. SAB Quarterly Sales (Millions of Liters)

Something that is not as apparent from the time series but that is pointed out by the periodogram is the presence of another cycle of period two quarters.

In constructing this table we set the interval equal to the sample size which is – after taking differences – 39 quarters. Thus if we have e.g. $i = 4$ the period is $\frac{39}{4}$ or roughly 10 quarters and the corresponding frequency is 0.1 waves per quarter. Note that we can not detect a cycle with a periodicity of say two weeks or even one full quarter as these cycles will not show up in the quarterly samples.

DEFINITION 1.40. In general, if the time between observations is defined as Δt then we cannot detect periods smaller than $2\Delta t$ or similarly frequencies higher than $\frac{\pi}{\Delta t}$. This limiting frequency is known as the *Nyquist frequency*.

Thus for the data in Figure 1.6 with $\Delta t = 1$ quarter the Nyquist frequency is π . Now π on the basic interval translates into half a quarter for this time series so that we require at least a frequency of half a wave per quarter or similarly a period of at least two quarters in order to be able to detect it.

i	Frequency	Period	Periodogram $I(f_i)$		
0	$\frac{0}{39} \approx 0.00$	$\frac{39}{0} \approx \infty$	113.70		
1	$\frac{1}{39} \approx 0.03$	$\frac{39}{1} \approx 39.00$	18.82	0.8%	•
2	$\frac{2}{39} \approx 0.05$	$\frac{39}{2} \approx 19.50$	2.90	0.1%	•
3	$\frac{3}{39} \approx 0.08$	$\frac{39}{3} \approx 13.00$	11.53	0.5%	•
4	$\frac{4}{39} \approx 0.10$	$\frac{39}{4} \approx 9.75$	3.74	0.2%	•
5	$\frac{5}{39} \approx 0.13$	$\frac{39}{5} \approx 7.80$	17.03	0.7%	•
6	$\frac{6}{39} \approx 0.15$	$\frac{39}{6} \approx 6.50$	8.67	0.4%	•
7	$\frac{7}{39} \approx 0.18$	$\frac{39}{7} \approx 5.57$	16.95	0.7%	•
8	$\frac{8}{39} \approx 0.21$	$\frac{39}{8} \approx 4.88$	74.07	3.2%	•
9	$\frac{9}{39} \approx 0.23$	$\frac{39}{9} \approx 4.33$	201.61	8.8%	•
10	$\frac{10}{39} \approx 0.26$	$\frac{39}{10} \approx 3.90$	1,201.34	52.5%	•
11	$\frac{11}{39} \approx 0.28$	$\frac{39}{11} \approx 3.55$	60.61	2.6%	•
12	$\frac{12}{39} \approx 0.31$	$\frac{39}{12} \approx 3.25$	43.26	1.9%	•
13	$\frac{13}{39} \approx 0.33$	$\frac{39}{13} \approx 3.00$	2.09	0.1%	•
14	$\frac{14}{39} \approx 0.36$	$\frac{39}{14} \approx 2.79$	0.90	0.0%	•
15	$\frac{15}{39} \approx 0.38$	$\frac{39}{15} \approx 2.60$	21.67	0.9%	•
16	$\frac{16}{39} \approx 0.41$	$\frac{39}{16} \approx 2.44$	21.59	0.9%	•
17	$\frac{17}{39} \approx 0.44$	$\frac{39}{17} \approx 2.29$	48.67	2.1%	•
18	$\frac{18}{39} \approx 0.46$	$\frac{39}{18} \approx 2.17$	65.69	2.9%	•
19	$\frac{19}{39} \approx 0.49$	$\frac{39}{19} \approx 2.05$	468.66	20.5%	•

TABLE 1.4. Periodogram of SAB Sales (First Differences)

We also need at least two points to specify a curve of the form (1.26) and the maximum frequency we can calculate from a sample of n data points is

$$n' = \begin{cases} \frac{n}{2}, & n \text{ even} \\ \frac{n-1}{2}, & n \text{ odd} \end{cases} \quad (1.43)$$

We cannot detect frequencies higher than 20 quarters from the 40 quarters of the SAB sales time series. After calculating first differences we are left

with 39 quarterly observations and we cannot detect frequencies higher than 19 quarters.

5.5. Fourier Series Representation of a Periodic Function. For simplicity we will assume that the functions we are dealing with in this section are all of period 2π as is the case with both $\cos$ and $\sin$. We can always change the period of a function to fit with this assumption – if $g(x)$ is periodic with period p then $f(x) = g(\frac{px}{2\pi})$ will also be periodic with period 2π .

We can represent any piecewise continuous periodic function as a sum of cosine waves through its *Fourier series representation*. This sum of cosine waves can also be written as the combined sum of cosine and sine waves as in (1.27).

THEOREM 1.41. *If $f(x)$ is periodic with period 2π with both $f(t)$ and $f'(t)$ piecewise continuous then $f(t)$ can be represented by its Fourier Series:*

$$f(t) = \frac{a_0}{2} + \sum_{j=1}^{\infty} [a_j \cos(jt) + b_j \sin(jt)] \quad (1.44)$$

$$= \sum_{j=1}^{\infty} d_j \cos(j\omega t + \theta_j) \quad (1.45)$$

with

$$a_j = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(jt) dt, \quad j = 0, 1, 2, \dots \quad (1.46)$$

$$b_j = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(jt) dt, \quad j = 1, 2, 3, \dots \quad (1.47)$$

and

$$\omega = \frac{2\pi}{p}. \quad (1.48)$$

PROOF. See for example Powell (1981:150-153). □

Here t is a continuous variable but we are working with a time series y_t observed at given intervals and do not have the luxury of a continuous function at our disposal. We have to approximate the integrals (1.46) and (1.47) and to do this we use the *discrete Fourier series approximation*.

THEOREM 1.42. *If only a finite and odd number of equally spaced function values*

$$f\left(\frac{2\pi t}{n}\right), \quad t = 0, 1, \dots, n-1$$

are known then the discrete Fourier series approximation of (1.46) and (1.47) is

$$a_j = \frac{2}{n} \sum_{i=0}^{n-1} f\left(\frac{2\pi i}{n}\right) \cos\left(\frac{2\pi j i}{n}\right), \quad j = 0, 1, \dots, n' \quad (1.49)$$

$$b_j = \frac{2}{n} \sum_{i=0}^{n-1} f\left(\frac{2\pi i}{n}\right) \sin\left(\frac{2\pi j i}{n}\right), \quad j = 1, 2, \dots, n' \quad (1.50)$$

with n' defined by (1.43).

PROOF. See e.g. Powell (1981:156-158). □

Note that if n is even it follows that $\frac{2\pi j i}{n} = i\pi$ if $j = n' - 1$ so that we have no $b_{n'}$ coefficient and we have to adjust $a_{n'}$ to half its original size.

The discrete Fourier series approximation of the time series is simply a general form of the linear cyclical model with mean $\frac{a_0}{2}$ and $m = n'$ different cycles P_k fitted to all possible periods $2, 3, \dots, n'$ present in the data. This approximation is exact but in practice we will normally want a parsimonious model consisting of only the important periods. Still the Fourier approximation provides a way of calculating the a_i and b_i parameters and of determining these important periods. The total variance of the Fourier approximation from (1.41) is

$$\begin{aligned} \sigma_{total}^2 &= \frac{1}{2} \sum_{i=1}^{n'} d_i^2 \\ &= \frac{1}{2} \sum_{i=1}^{n'} (a_i + b_i)^2 \\ &= 2 \sum_{i=1}^{n'} I\left(\frac{2\pi i}{n}\right) \end{aligned}$$

if n is odd. The approximation is exact and we no longer need ε_t in the expression of y_t and therefore σ_ε^2 has fallen away. If n is even then we have

only one degree of freedom for the last term as $b_{n'}$ falls away and we calculate

$$I(\pi) = \frac{a_{n'}^2}{2}.$$

The contribution to the total variance of any one wave i is obtained from the intensity at that frequency. As such the Fourier approximation componentised the variance according to frequency (Rayner 1971:23) or, put differently, transformed the series from the time to the frequency domain. Using (1.41) we can construct any autocovariance and any autocorrelation from which we obtain the autocovariance generating function.

If we for example use only one wave (of period 4 quarters) to model the SAB quarterly sales we will from Table 1.4 be able to explain 53% of the variance in this data. This approach is not without its drawbacks and even Schuster noted that the results are sometimes misleading (Rayner 1971:25). Meyers (April 1999:16) noted that although the *in sample* fit is excellent a Fourier series approximation provides misleading and inaccurate forecasts. Still he recommended it as a trend determinant (May 1999:26).

As a final example consider the percentage returns of both the JSE Overall Index and the ZAR/USD exchange rate as in Figure 1.7. These series appear periodic but note how difficult it is to obtain the periods from only inspecting this figure. Tables 1.5 and 1.6 make the task very easy (only the first few rows of each table is displayed – with the exception of a 9 month period for both time series, none of the higher frequencies were significant). It is interesting to note that both these series contain similar cycles and both periodograms have peaks at around two and a half year cycles. Cyclical behaviour in asset returns is well known and generally becomes evident as the frequency at which observations are made is lowered – normally to monthly observations.

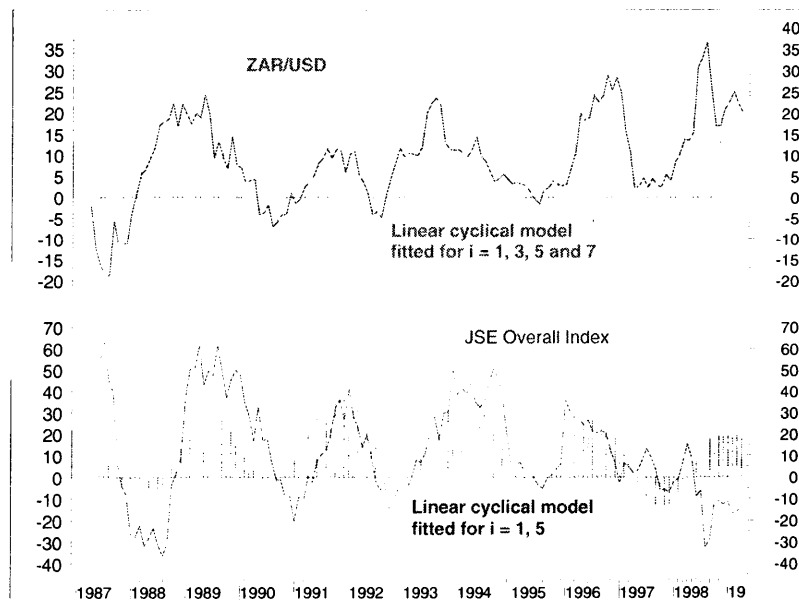


FIGURE 1.7. Percentage Returns (Month on Year)

i	Frequency	Period	Periodogram $I(f_i)$			
0	$\frac{0}{143} \approx 0.00$	$\frac{143}{0} \approx \infty$	152.30			
1	$\frac{1}{143} \approx 0.01$	$\frac{143}{1} \approx 143.00$	18.19	6.8%	—•	
2	$\frac{2}{143} \approx 0.01$	$\frac{143}{2} \approx 71.50$	30.51	11.3%	—•	
3	$\frac{3}{143} \approx 0.02$	$\frac{143}{3} \approx 47.67$	21.53	8.0%	—•	
4	$\frac{4}{143} \approx 0.03$	$\frac{143}{4} \approx 35.75$	1.65	0.6%	•	
5	$\frac{5}{143} \approx 0.03$	$\frac{143}{5} \approx 28.60$	90.69	33.7%	—•	
6	$\frac{6}{143} \approx 0.04$	$\frac{143}{6} \approx 23.83$	22.50	8.4%	—•	
7	$\frac{7}{143} \approx 0.05$	$\frac{143}{7} \approx 20.43$	11.83	4.4%	—•	
8	$\frac{8}{143} \approx 0.06$	$\frac{143}{8} \approx 17.88$	23.62	8.8%	—•	
9	$\frac{9}{143} \approx 0.06$	$\frac{143}{9} \approx 15.89$	5.37	2.0%	—•	
10	$\frac{10}{143} \approx 0.07$	$\frac{143}{10} \approx 14.30$	3.64	1.4%	—•	

TABLE 1.5. JSE Overall Index Percentage Returns (Month on Year)

i	Frequency	Period	Periodogram $I(f_i)$		
0	$\frac{0}{143} \approx 0.00$	$\frac{143}{0} \approx \infty$	81.97		
1	$\frac{1}{143} \approx 0.01$	$\frac{143}{1} \approx 143.00$	2.82	5.4%	—•
2	$\frac{2}{143} \approx 0.01$	$\frac{143}{2} \approx 71.50$	0.17	0.3%	•
3	$\frac{3}{143} \approx 0.02$	$\frac{143}{3} \approx 47.67$	7.24	13.7%	———•
4	$\frac{4}{143} \approx 0.03$	$\frac{143}{4} \approx 35.75$	0.63	1.2%	—•
5	$\frac{5}{143} \approx 0.03$	$\frac{143}{5} \approx 28.60$	21.04	39.9%	—————•
6	$\frac{6}{143} \approx 0.04$	$\frac{143}{6} \approx 23.83$	4.62	8.8%	———•
7	$\frac{7}{143} \approx 0.05$	$\frac{143}{7} \approx 20.43$	7.03	13.3%	———•
8	$\frac{8}{143} \approx 0.06$	$\frac{143}{8} \approx 17.88$	0.22	0.4%	•
9	$\frac{9}{143} \approx 0.06$	$\frac{143}{9} \approx 15.89$	0.06	0.1%	•
10	$\frac{10}{143} \approx 0.07$	$\frac{143}{10} \approx 14.30$	0.52	1.0%	—•

TABLE 1.6. ZAR/USD Percentage Returns (Month on Year)

6. Difference Equations

6.1. First order difference equations. A difference equation is a linear combination that relates a time series to its previous values, such as the *linear first order difference equation*.

DEFINITION 1.43. The linear first order difference equation expresses a time series y_t as a function of its previous occurrence y_{t-1} and some other time series variable ε_t

$$y_t = \phi y_{t-1} + \varepsilon_t. \quad (1.51)$$

This ‘other’ variable ε_t is not necessarily white noise but it will often be as hinted by the notation adopted. In practice it could even be some combination of deterministic variables and white noise in the form of an econometric model

$$\varepsilon_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_p x_{p,t} + e_t$$

with $x_{i,j}$ econometric time series such as interest and exchange rates although we are not too concerned about that here.

Difference equations are often solved through a process of *recursive back substitution*:

$$\begin{aligned} y_t &= \phi y_{t-1} + \varepsilon_t \\ &= \phi(\phi y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \phi^2(\phi y_{t-3} + \varepsilon_{t-2}) + \phi \varepsilon_{t-1} + \varepsilon_t \\ &\vdots \\ &= \phi^t y_0 + \phi^{t-1} \varepsilon_1 + \phi^{t-2} \varepsilon_2 + \cdots + \phi \varepsilon_{t-1} + \varepsilon_t. \end{aligned}$$

Here we assume that y_t starts initially with value y_0 but the substitution can be stopped earlier or carried on longer depending on where the series has its initial value. If we carry on indefinitely we end up with an infinite linear combination of the ε_{t-i} with weights ϕ^i .

THEOREM 1.44. *If ε_t is white noise then it follows immediately from Corollary 1.15 that the first order difference equation (1.51) is stationary if, and only if,*

$$|\phi| < 1.$$

6.1.1. Dynamic Multipliers.

DEFINITION 1.45. The effect of ε_{t-k} on y_t is known as the *dynamic multiplier*

$$\frac{\partial y_t}{\partial \varepsilon_{t-k}}. \quad (1.52)$$

If we back substitute far enough we can easily calculate the dynamic multiplier for (1.51) as

$$\frac{\partial y_t}{\partial \varepsilon_{t-k}} = \phi^k. \quad (1.53)$$

If $|\phi| < 1$ this multiplier decays geometrically to zero and leads to a stable, stationary system as noted in Theorem 1.44. The problem is if $|\phi| > 1$ in which case the system explodes since the dynamic multiplier increases

exponentially. If $|\phi| = 1$ we have an interesting borderline case, called a *random walk*, that we will examine later on.

6.2. Higher Order Difference Equations.

DEFINITION 1.46. The *difference equation of order p* is defined as

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \cdots + \phi_p y_{t-p} + \varepsilon_t. \quad (1.54)$$

Here we allowed more than just the previous occurrence to enter the equation. We can rewrite (1.54) conveniently as a *first order vector difference equation*

$$\begin{pmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-(p-1)} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ y_{t-3} \\ \vdots \\ y_{t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

or

$$\xi_t = \mathbf{F} \xi_{t-1} + \varepsilon_t \quad (1.55)$$

The first equation of this system of p equations is (1.54) while the next equation is simply $y_{t-1} = y_{t-1}$, the third equation is $y_{t-2} = y_{t-2}$ and so on until the final equation yields $y_{t-(p-1)} = y_{t-(p-1)}$.

We can apply back substitution to (1.55) as before to obtain

$$\xi_t = \mathbf{F}^t \xi_0 + \mathbf{F}^{t-1} \varepsilon_1 + \mathbf{F}^{t-2} \varepsilon_2 + \cdots + \mathbf{F} \varepsilon_{t-1} + \varepsilon_t. \quad (1.56)$$

If we define $f_{ij}^{(t)}$ as the (i, j) th element of the matrix $\mathbf{F}^t$ then from (1.56)

$$\begin{aligned} y_t &= f_{11}^{(t)} y_0 + f_{12}^{(t)} y_{-1} + f_{13}^{(t)} y_{-2} + \cdots + f_{1p}^{(t)} y_{-(p-1)} \\ &\quad + f_{11}^{(t-1)} \varepsilon_1 + f_{11}^{(t-2)} \varepsilon_2 + f_{11}^{(t-3)} \varepsilon_3 + \cdots + f_{11}^{(1)} \varepsilon_{t-1} + \varepsilon_t. \end{aligned} \quad (1.57)$$

This linear function requires p initial values of y_t as opposed to the single y_0 required by (1.51). As before we do not need to stop the substitution at this point but can carry on indefinitely in order to yield an infinite linear combination in ε_t .

6.2.1. *Dynamic Multipliers.* It follows immediately from (1.57) that the dynamic multiplier for a higher order difference equation is

$$\frac{\partial y_t}{\partial \varepsilon_{t-j}} = f_{11}^{(j)}. \quad (1.58)$$

If $j = 1$ the dynamic multiplier simply is ϕ_1 . We can quickly find the dynamic multiplier for larger j through numerically simulating the system but it is also useful to have an analytical expression for it. Since it is given by the $(1, 1)$ element of $\mathbf{F}^j$ we need to investigate this matrix further. We are specifically interested in the eigenvalues of $\mathbf{F}$ which dictates the behaviour of $\mathbf{F}^j$ as $j \rightarrow \infty$.

6.2.2. *Distinct Eigenvalues.* If the eigenvalues λ_i of $\mathbf{F}$ are distinct it is well known (Fraleigh & Beauregard 1987:230) that we can diagonalize the matrix $\mathbf{F}$ as

$$\mathbf{F}^j = \mathbf{M} \cdot \mathbf{\Lambda}^j \cdot \mathbf{M}^{-1} \quad (1.59)$$

where

$$\mathbf{\Lambda}^j = \begin{pmatrix} \lambda_1^j & 0 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2^j & 0 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3^j & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & \lambda_p^j \end{pmatrix}. \quad (1.60)$$

If m_{ij} is the (i, j) element of $\mathbf{M}$ and similarly $m_{ij}^{(-1)}$ is the (i, j) element of $\mathbf{M}^{-1}$ it follows that the $(1, 1)$ element of $\mathbf{F}^j$ is

$$\begin{aligned} f_{11}^{(j)} &= m_{11} m_{11}^{(-1)} \lambda_1^j + m_{12} m_{21}^{(-1)} \lambda_2^j + m_{13} m_{31}^{(-1)} \lambda_3^j + \cdots + m_{1p} m_{p1}^{(-1)} \lambda_p^j \\ &= c_1 \lambda_1^j + c_2 \lambda_2^j + c_3 \lambda_3^j + \cdots + c_p \lambda_p^j \\ &= \frac{\partial y_t}{\partial \varepsilon_{t-j}}. \end{aligned} \quad (1.61)$$

Since $c_i = m_{1i} \cdot m_{i1}^{(-1)}$ the sum of the c_i is

$$\begin{aligned} \sum_{i=1}^p c_i &= \sum_{i=1}^p m_{1i} m_{i1}^{(-1)} \\ &= \text{first element of } \mathbf{M} \cdot \mathbf{M}^{-1} \text{ which is } \mathcal{I}_p \\ &= 1. \end{aligned}$$

Thus the dynamic multiplier $f_{11}^{(j)}$ is a weighted average of the eigenvalues of $\mathbf{F}$ raised to the power j . Since the dynamic multiplier is determined by these c_i they are quite important. We can actually calculate them explicitly as the elements of a *Vandermonde's* matrix.

6.2.3. *Vandermonde's Matrices.* Alexander-Theophile Vandermonde (1735-1796) in 1771 gave various methods for computing determinants (Fraleigh & Beauregard 1987:200). These methods apply especially to matrices containing eigenvectors as in our present situation.

DEFINITION 1.47. A Vandermonde matrix is defined as

$$a_{ij} = x_j^i \quad (1.62)$$

thus

$$\mathbf{A} = \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ x_1^2 & x_2^2 & x_3^2 & \cdots & x_n^2 \\ x_1^3 & x_2^3 & x_3^3 & \cdots & x_n^3 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_1^n & x_2^n & x_3^n & \cdots & x_n^n \end{pmatrix} \quad (1.63)$$

THEOREM 1.48. *The determinant of a Vandermonde's matrix is*

$$\prod_{j=1}^n x_j \prod_{i=j+1}^n (x_i - x_j). \quad (1.64)$$

PROOF. Following Knuth's suggestions (1973:35) we subtract the first column from the others

$$\begin{pmatrix} x_1 & x_2 - x_1 & x_3 - x_1 & \cdots & x_n - x_1 \\ x_1^2 & x_2^2 - x_1^2 & x_3^2 - x_1^2 & \cdots & x_n^2 - x_1^2 \\ x_1^3 & x_2^3 - x_1^3 & x_3^3 - x_1^3 & \cdots & x_n^3 - x_1^3 \\ \dots & \dots & \dots & \dots & \dots \\ x_1^n & x_2^n - x_1^n & x_3^n - x_1^n & \cdots & x_n^n - x_1^n \end{pmatrix}$$

and then we subtract x_1 times row $k-1$ from row k for $k = n, n-1, n-2, \dots, 2$ to obtain (after some simplification)

$$\begin{pmatrix} x_1 & x_2 - x_1 & x_3 - x_1 & \cdots & x_n - x_1 \\ 0 & x_2(x_2 - x_1) & x_3(x_3 - x_1) & \cdots & x_n(x_n - x_1) \\ 0 & x_2^2(x_2 - x_1) & x_3^2(x_3 - x_1) & \cdots & x_n^2(x_n - x_1) \\ \dots & \dots & \dots & \dots & \dots \\ 0 & x_2^{n-1}(x_2 - x_1) & x_3^{n-1}(x_3 - x_1) & \cdots & x_n^{n-1}(x_n - x_1) \end{pmatrix}.$$

We now factor out x_1 from the first column and $x_j - x_1$ from the rest to end up with $x_1(x_2 - x_1)(x_3 - x_1) \dots (x_n - x_1)$ times a Vandermonde determinant of order $n-1$. The result then follows by recursion. $\square$

LEMMA 1.49. *The coefficient c_{ij} of x^{j-1} and c_{ij}^* of x^{n-j} in the expression*

$$\prod_{\substack{k=1 \\ k \neq i}}^n (x_k - x) \quad (1.65)$$

is given by

$$c_{ij} = (-1)^{j-1} \sum_{\substack{1 \leq k_1 < k_2 < \dots < k_{n-j} \leq n \\ k_1, k_2, \dots, k_{n-j} \neq i}} x_{k_1} x_{k_2} \dots x_{k_{n-j}} \quad (1.66)$$

$$c_{ij}^* = (-1)^{n-j} \sum_{\substack{1 \leq k_1 < k_2 < \dots < k_{j-1} \leq n \\ k_1, k_2, \dots, k_{j-1} \neq i}} x_{k_1} x_{k_2} \dots x_{k_{j-1}} \quad (1.67)$$

PROOF. We can write (1.65) as

$$\prod_{\substack{k=1 \\ k \neq i}}^n (x_k - x) = (x_1 - x)(x_2 - x) \dots (x_{i-1} - x)(x_{i+1} - x) \dots (x_n - x)$$

and then the sum of all possible combinations of $n - j$ distinct x_k 's gives the required coefficient c_{ij} with the necessary adjustment for its sign determined by the amount $j - 1$ of $-x$'s and similarly for c_{ij}^* . $\square$

COROLLARY 1.50. *The inverse $\mathbf{B}$ of a Vandermonde's matrix $\mathbf{A}$ is*

$$b_{ij} = (-1)^{j+1} \cdot \frac{\sum_{\substack{1 \leq k_1 < k_2 < \dots < k_{n-j} \leq n \\ k_1, k_2, \dots, k_{n-j} \neq i}} x_{k_1} x_{k_2} \dots x_{k_{n-j}}}{x_i \prod_{\substack{k=1 \\ k \neq i}}^n (x_k - x_i)}. \quad (1.68)$$

PROOF. First note that the sum in the numerator is the c_{ij} coefficient of (1.65) and then it follows that

$$\begin{aligned} \sum_{k=1}^n b_{ij} a_{jk} &= \sum_{k=1}^n \frac{c_{ij} x_k^j}{x_i \prod_{\substack{t=1 \\ t \neq i}}^n (x_t - x_i)} \\ &= \frac{x_k \prod_{\substack{t=1 \\ t \neq i}}^n (x_t - x_k)}{x_i \prod_{\substack{t=1 \\ t \neq i}}^n (x_t - x_i)} \\ &= \begin{cases} 1, & k = i \\ 0, & k \neq i \end{cases}. \end{aligned}$$

$\square$

In the time series context we rather work with a matrix that has the format

$$a_{ij}^* = x_j^{n-i+1}$$

or

$$\mathbf{A}^* = \begin{pmatrix} x_1^n & x_2^n & \dots & x_n^n \\ x_1^{n-1} & x_2^{n-1} & \dots & x_n^{n-1} \\ \dots & \dots & \dots & \dots \\ x_1 & x_2 & \dots & x_n \end{pmatrix}$$

which we will also refer to as a Vandermonde's matrix. Note that the determinant of this matrix is

$$|\mathbf{A}^*| = -|\mathbf{A}| = -\prod_{j=1}^n x_j \prod_{i=j+1}^n (x_j - x_i)$$

with $\mathbf{A}$ a regular Vandermonde's matrix $a_{ij} = x_j^i$. The inverse of this matrix can be derived using c_{ij}^* (1.67) as

$$b_{ij}^* = (-1)^{n-j} \cdot \frac{\sum_{\substack{1 \leq k_1 < k_2 < \dots < k_{j-1} \leq n \\ k_1, k_2, \dots, k_{j-1} \neq i}} x_{k_1} x_{k_2} \dots x_{k_{j-1}}}{x_i \prod_{\substack{k=1 \\ k \neq i}}^n (x_k - x_i)}. \quad (1.69)$$

We will also make use of the well known result that if a real matrix has a complex eigenvalue, say

$$\lambda_i = \alpha + \beta\iota \quad \text{with } \iota = \sqrt{-1}$$

then the complex conjugate of λ_i , say

$$\lambda_j = \alpha - \beta\iota.$$

is also an eigenvalue.

In order to apply these results to difference equations we use the next theorem.

THEOREM 1.51. *The matrix $\mathbf{M}$ containing the eigenvectors of $\mathbf{F}$ in (1.60) is a Vandermonde's matrix.*

PROOF. If λ_i is an eigenvalue of $\mathbf{F}$ it satisfies

$$\lambda_i^p - \theta_1 \lambda_i^{p-1} - \theta_2 \lambda_i^{p-2} - \dots - \theta_{p-1} \lambda_i - \theta_p = 0 \quad (1.70)$$

as shown later on (1.83). If we construct a Vandermonde's matrix $\mathbf{M}$ in a step by step manner with i th column

$$\mathbf{m}_i = \begin{pmatrix} \lambda_i^p \\ \lambda_i^{p-1} \\ \vdots \\ \lambda_i \end{pmatrix}$$

we have

$$\begin{aligned} \mathbf{F} \cdot \mathbf{m}_i &= \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \lambda_i^p \\ \lambda_i^{p-1} \\ \lambda_i^{p-2} \\ \vdots \\ \lambda_i \end{pmatrix} \\ &= \begin{pmatrix} \phi_1 \lambda_i^p + \phi_2 \lambda_i^{p-1} + \cdots + \phi_p \lambda_i \\ \lambda_i^p \\ \lambda_i^{p-1} \\ \vdots \\ \lambda_i^2 \end{pmatrix} \end{aligned}$$

and, substituting (1.70) into the first element yields

$$\begin{aligned} &= \begin{pmatrix} \lambda_i^{p+1} \\ \lambda_i^p \\ \lambda_i^{p-1} \\ \vdots \\ \lambda_i^2 \end{pmatrix} \\ &= \lambda_i \mathbf{m}_i \end{aligned}$$

which shows that $\mathbf{m}_i$ is an eigenvector of $\mathbf{F}$ corresponding to eigenvalue λ_i . $\square$

COROLLARY 1.52. *If the eigenvalues λ_i of the matrix $\mathbf{F}$ are distinct then c_i of (1.61) is*

$$c_i = \frac{\lambda_i^{p-1}}{\prod_{\substack{j=1 \\ j \neq i}}^p (\lambda_i - \lambda_j)}. \quad (1.71)$$

PROOF. We can write c_i as in (1.61) as

$$c_i = m_{1i} m_{i1}^{(-1)} = a_{1i}^* b_{i1}^*$$

and, substituting (1.69) into this with $x_i = \lambda_i$ yields

$$c_i = \frac{\lambda_i^n}{\lambda_i \prod_{\substack{t=1 \\ t \neq i}}^n (\lambda_t - \lambda_i)}$$

which is the result. $\square$

COROLLARY 1.53. *If λ_i, λ_j are complex conjugates and eigenvalues of the real matrix $\mathbf{F}$ then c_i and c_j are also complex conjugates.*

PROOF. This follows directly from substituting complex conjugates into (1.71). $\square$

Since both λ_i and λ_j as well as c_i and c_j are complex conjugates it turns out – quite helpfully – that (1.61) will always be real for real $\mathbf{F}$.

6.2.4. Repeated Eigenvalues. In the even more general case where some of the eigenvalues of $\mathbf{F}$ are repeated so that we have $s < p$ linearly independent eigenvectors we make use of Camille Jordan's *Jordan canonical form* (Fraleigh & Beauregard 1987:243)

$$\mathbf{F} = \mathbf{M} \cdot \mathbf{J} \cdot \mathbf{M}^{-1}$$

with $\mathbf{J}$ constituting of the matrices $\mathbf{J}_i$ as diagonal 'elements'

$$\mathbf{J} = \begin{pmatrix} \mathbf{J}_1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_2 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{J}_3 & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{J}_s \end{pmatrix}$$

where $\mathbf{J}_i$ is the matrix containing as diagonal elements all occurrences of the repeated eigenvalue λ_i along with ones in the next diagonal

$$\mathbf{J}_i = \begin{pmatrix} \lambda_i & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 & 0 \\ 0 & 0 & \lambda_i & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \lambda_i & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda_i \end{pmatrix}$$

so that, as before,

$$\mathbf{F}^j = \mathbf{M} \cdot \mathbf{J}^j \cdot \mathbf{M}^{-1}.$$

THEOREM 1.54. *If $\mathbf{J}$ is an $n \times n$ matrix*

$$\mathbf{J} = \begin{pmatrix} \lambda & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \lambda & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & \lambda \end{pmatrix}$$

then

$$\mathbf{J}^j = \begin{pmatrix} \lambda^j & \binom{j}{1}\lambda^{j-1} & \binom{j}{2}\lambda^{j-2} & \cdots & \binom{j}{n-1}\lambda^{j-n+1} \\ 0 & \lambda^j & \binom{j}{1}\lambda^{j-1} & \cdots & \binom{j}{n-2}\lambda^{j-n+2} \\ 0 & 0 & \lambda^j & \cdots & \binom{j}{n-3}\lambda^{j-n+3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \lambda^j \end{pmatrix} \quad \text{for } j = 1, 2, \dots$$

PROOF. This is shown using mathematical induction and making use of the relations (Hamilton 1994:19)

$$\binom{j}{n} = \begin{cases} \frac{j(j-1)(j-2)\cdots(j-n+1)}{n(n-1)(n-2)\cdots 3 \cdot 2 \cdot 1}, & j \geq n \\ 0, & j < n \end{cases} \quad (1.72)$$

and

$$\binom{j+1}{n} = \binom{j}{n} + \binom{j}{n-1}. \quad (1.73)$$

The theorem holds for $i = 1$, mainly through (1.72), and, if it holds for $j = k$ then we have

$$\begin{aligned} \mathbf{J}^{k+1} &= \mathbf{J}^k \cdot \mathbf{J} \\ &= \begin{pmatrix} \lambda^k & \binom{k}{1}\lambda^{k-1} & \binom{k}{2}\lambda^{k-2} & \dots & \binom{k}{n-1}\lambda^{k-n+1} \\ 0 & \lambda^k & \binom{k}{1}\lambda^{k-1} & \dots & \binom{k}{n-2}\lambda^{k-n+2} \\ 0 & 0 & \lambda^k & \dots & \binom{k}{n-3}\lambda^{k-n+3} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \lambda^k \end{pmatrix} \\ &\quad \times \begin{pmatrix} \lambda & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & \lambda & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \lambda & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & \lambda \end{pmatrix} \end{aligned}$$

which, from (1.73), becomes

$$= \begin{pmatrix} \lambda^{k+1} & \binom{k+1}{1}\lambda^k & \binom{k+1}{2}\lambda^{k-1} & \dots & \binom{k+1}{n-1}\lambda^{k+1-n+1} \\ 0 & \lambda^{k+1} & \binom{k+1}{1}\lambda^k & \dots & \binom{k+1}{n-2}\lambda^{k+1-n+2} \\ 0 & 0 & \lambda^{k+1} & \dots & \binom{k+1}{n-3}\lambda^{k+1-n+3} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \lambda^{k+1} \end{pmatrix}.$$

□

For a second order difference equation with a repeated eigenvalue λ we will e.g. have

$$\mathbf{F}^j = \mathbf{M} \begin{pmatrix} \lambda^j & j\lambda^{j-1} \\ 0 & \lambda^j \end{pmatrix} \mathbf{M}^{-1}.$$

The dynamic multiplier of this system is

$$\begin{aligned} f_{11}^{(j)} &= m_{11}m_{11}^{(-1)}\lambda^j + m_{12}m_{21}^{(-1)}j\lambda^{j-1} \\ &= c_1\lambda^j + c_2j\lambda^{j-1}. \end{aligned} \tag{1.74}$$

6.3. Stable Systems.

THEOREM 1.55. *Only if all eigenvalues of a matrix $\mathbf{F}$ lie within the unit circle, $|\lambda| < 1$, then*

$$\sum_{j=0}^{\infty} \mathbf{F}^j = (\mathcal{I} - \mathbf{F})^{-1}. \quad (1.75)$$

PROOF. This is the matrix equivalent of the geometric series and we do not provide a complete proof.

Firstly, suppose that the inverse $(\mathcal{I} - \mathbf{F})^{-1}$ does not exist. Then the determinant is zero

$$|\mathcal{I} - \mathbf{F}| = |\mathbf{F} - \mathcal{I}| = 0$$

and 1 is an eigenvalue of $\mathbf{F}$ which is not strictly within the unit circle. Thus the inverse exists.

It can easily be shown that

$$\mathbf{F}^j = \mathbf{M} \Lambda^j \mathbf{M}^{-1} \rightarrow \mathbf{0} \quad \text{as } j \rightarrow \infty$$

only if all the eigenvalues are within the unit circle. We now calculate

$$\begin{aligned} (\mathcal{I} - \mathbf{F})(\mathcal{I} + \mathbf{F} + \mathbf{F}^2 + \dots + \mathbf{F}^j) &= (\mathcal{I} + \mathbf{F} + \mathbf{F}^2 + \dots + \mathbf{F}^j)(\mathcal{I} - \mathbf{F}) \\ &= \mathcal{I} - \mathbf{F}^{j+1} \\ &\rightarrow \mathcal{I} \quad \text{as } j \rightarrow \infty. \end{aligned}$$

Thus

$$(\mathcal{I} - \mathbf{F})(\mathcal{I} + \mathbf{F} + \mathbf{F}^2 + \dots) = (\mathcal{I} + \mathbf{F} + \mathbf{F}^2 + \dots)(\mathcal{I} - \mathbf{F}) = \mathcal{I}$$

showing that $(\mathcal{I} - \mathbf{F})^{-1} = (\mathcal{I} + \mathbf{F} + \mathbf{F}^2 + \dots)$. $\square$

If the eigenvalues of $\mathbf{F}$ are all less than 1 in magnitude then $\mathbf{F}^j$ decays to zero as j increases and the system is *stable*. Then we can think of y_t in terms of the infinite history of ε_t

$$y_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots$$

where $\psi_j = f_{11}^{(j)}$ is simply the first element of $\mathbf{F}^j$. This infinite linear combination of ε_t allows us, in terms of the results of Section 2 to investigate the stationarity of the system in terms of the ψ_i weights. We can also calculate

the *full effect* of a change in ε_t on y_t . The cumulative or full effect of a change in ε_t on $\xi_t, \xi_{t+1}, \dots$ is given by

$$\begin{aligned} \frac{\partial \sum_{i=0}^{\infty} \xi_{t+i}}{\partial \varepsilon_t} &= \sum_{i=0}^{\infty} \mathbf{F}^i \\ &= (\mathbf{I} - \mathbf{F})^{-1} \end{aligned} \quad (1.76)$$

where (1.76) exists provided that the eigenvalues of $\mathbf{F}$ are all within the unit circle. The first element of (1.76) is

$$\frac{\partial \sum_{i=0}^{\infty} y_{t+i}}{\partial \varepsilon_t} \quad (1.77)$$

and gives the cumulative effect of a change in ε_t on $y_t, y_{t+1}, \dots$. This is also called the *long run effect* of a change in ε_t on y_t (Hamilton 1994:23).

THEOREM 1.56. *If the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_p$ of the $p \times p$ matrix $\mathbf{F}$ lie within the unit circle the matrix*

$$(\mathbf{I}_p - \mathbf{F})^{-1}$$

exists with (1,1) element

$$\frac{1}{1 - \phi_1 - \phi_2 - \dots - \phi_p}.$$

PROOF. We know that the matrix exists by Theorem 1.55. In order to solve for the (1,1) element of this matrix we write, with x_{ij} the (i,j) th element of $(\mathbf{I}_p - \mathbf{F})^{-1}$

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \dots & \dots & \dots & \dots \\ x_{p1} & x_{p2} & \cdots & x_{pp} \end{pmatrix} \times \begin{pmatrix} 1 - \phi_1 & -\phi_2 & \cdots & -\phi_{p-1} & -\phi_p \\ -1 & 1 & \cdots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \cdots & -1 & 1 \end{pmatrix} = \mathbf{I}_p \quad (1.78)$$

□

DEFINITION 1.57. The characteristic equation of the difference equation (1.54) is given by

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \phi_3 z^3 - \cdots - \phi_p z^p. \quad (1.80)$$

The next theorem allows us to translate finding the eigenvalues of the matrix $\mathbf{F}$ into finding the roots of this polynomial. If we substitute z in the characteristic equation (1.80) with the backward shift operator B we can rewrite the difference equation quite usefully as

$$\begin{aligned} y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \cdots - \phi_p y_{t-p} &= \varepsilon_t \\ (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) y_t &= \varepsilon_t \\ \phi(B) y_t &= \varepsilon_t. \end{aligned} \quad (1.81)$$

We now proof the main result after Hamilton (1994:21).

THEOREM 1.58. *Factoring the p th order polynomial or characteristic equation*

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = (1 - \lambda_1 z)(1 - \lambda_2 z) \cdots (1 - \lambda_p z)$$

is the same as finding the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_p$ of the matrix

$$\mathbf{F} = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

PROOF. The eigenvalues of $\mathbf{F}$ satisfy $|\mathbf{F} - \lambda \mathbf{I}| = 0$ so that

$$\begin{vmatrix} \phi_1 - \lambda & \phi_2 & \phi_3 & \cdots & \phi_{p-1} & \phi_p \\ 1 & -\lambda & 0 & \cdots & 0 & 0 \\ 0 & 1 & -\lambda & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & -\lambda \end{vmatrix} = 0. \quad (1.82)$$

If we multiply the p th column with $\frac{1}{\lambda}$ and add this to the $(p-1)$ th column we get

$$\begin{vmatrix} \phi_1 - \lambda & \phi_2 & \phi_3 & \cdots & \phi_{p-1} + \frac{\phi_p}{\lambda} & \phi_p \\ 1 & -\lambda & 0 & \cdots & 0 & 0 \\ 0 & 1 & -\lambda & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & -\lambda & 0 \\ 0 & 0 & 0 & \cdots & 0 & -\lambda \end{vmatrix} = 0.$$

Continuing in this fashion reduces the matrix to an upper triangular matrix

$$\begin{vmatrix} \sum_{i=1}^p \frac{\phi_i}{\lambda^{i-1}} - \lambda & \sum_{i=2}^p \frac{\phi_i}{\lambda^{i-1}} & \sum_{i=3}^p \frac{\phi_i}{\lambda^{i-1}} & \cdots & \phi_{p-1} + \frac{\phi_p}{\lambda} & \phi_p \\ 0 & -\lambda & 0 & \cdots & 0 & 0 \\ 0 & 0 & -\lambda & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 & -\lambda \end{vmatrix} = 0.$$

The determinant of this upper triangular matrix is the product of its diagonal elements

$$\begin{aligned} |\mathbf{F} - \lambda \mathbf{I}| &= \left(\left[\sum_{i=1}^p \frac{\phi_i}{\lambda^{i-1}} \right] - \lambda \right) (-\lambda)^{p-1} \\ &= (-1)^{p-1} \lambda^{p-1} \left(\phi_1 + \frac{\phi_2}{\lambda} + \cdots + \frac{\phi_p}{\lambda^{p-1}} - \lambda \right) \\ &= (-1)^p (\lambda^p - \phi_1 \lambda^{p-1} - \phi_2 \lambda^{p-2} - \cdots - \phi_p) \end{aligned} \quad (1.83)$$

which is also the determinant of (1.82).

On the other hand, we can write the polynomial as

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = (1 - \lambda_1 z)(1 - \lambda_2 z) \cdots (1 - \lambda_p z)$$

and, setting $\lambda = \frac{1}{z}$ and dividing both sides by z^p yields

$$\lambda^p - \phi_1 \lambda^{p-1} - \phi_2 \lambda^{p-2} - \cdots - \phi_{p-1} \lambda - \phi_p = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_p).$$

This expression has its roots at $\lambda = \lambda_i$ which are also the eigenvalues of the matrix $\mathbf{F}$ from (1.83). $\square$

COROLLARY 1.59. *Requiring that the eigenvalues of the matrix $\mathbf{F}$ lie inside the unit circle is the same as requiring that the roots of the characteristic equation lie outside the unit circle (and vice versa).*

PROOF. The roots of the characteristic equation can be obtained from $1 - \lambda_i z = 0$ or $z = \frac{1}{\lambda_i}$ so that $|z| > 1$ only if $|\lambda_i| < 1$. $\square$

COROLLARY 1.60. *The difference equation (1.54) is stable if the eigenvalues of the matrix $\mathbf{F}$ are inside, or, similarly, the roots of the characteristic equation are outside the unit circle.*

Note that the above holds, even if some of the eigenvalues are the same.

7. General Linear Models

We now have two separate but equivalent general linear models available to represent the stationary time series y_t which for the remainder of this section we will assume have been transformed to a zero mean time series z_t . It is a trivial matter to transform a stationary time series to a zero mean stationary time series through

$$z_t = y_t - \mu. \quad (1.84)$$

Firstly we can model our time series as a difference equation on its own past z_{t-i} and one current error ε_t

$$\begin{aligned} z_t &= \varepsilon_t + \pi_1 z_{t-1} + \pi_2 z_{t-2} + \dots \\ &= \varepsilon_t + \sum_{i=1}^{\infty} \pi_i z_{t-i}. \end{aligned}$$

Thus we may write

$$\begin{aligned} \varepsilon_t &= z_t - \sum_{i=1}^{\infty} \pi_i z_{t-i} \\ &= \left[1 - \sum_{i=1}^{\infty} \pi_i B^i \right] z_t \end{aligned}$$

If we define

$$\begin{aligned}\pi(z) &= 1 - \pi_1 z - \pi_2 z^2 - \dots \\ &= 1 - \sum_{i=1}^{\infty} \pi_i z^i\end{aligned}\tag{1.85}$$

we may write, using the backward shift operator

$$\varepsilon_t = \pi(B)z_t.\tag{1.86}$$

We can also, under suitable conditions, equivalently model our time series as a weighted sum of ε_t

$$\begin{aligned}z_t &= \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots \\ &= \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}\end{aligned}$$

with $\psi_0 = 1$ and, if we define

$$\psi(z) = \sum_{i=0}^{\infty} \psi_i z^i\tag{1.87}$$

we can write

$$\begin{aligned}z_t &= \left[\sum_{i=0}^{\infty} \psi_i B^i \right] \varepsilon_t \\ &= \psi(B)\varepsilon_t.\end{aligned}\tag{1.88}$$

The white noise ε_t may be considered as a series of shocks that drive the system (Box & Jenkins 1976:46).

From (1.88) and (1.86) it follows then that

$$\begin{aligned}\varepsilon_t &= \pi(B)z_t \\ &= \pi(B)\psi(B)\varepsilon_t\end{aligned}$$

so that

$$\pi(B)\psi(B) = 1$$

or

$$\pi(B) = \psi^{-1}(B). \quad (1.89)$$

We can use this relationship to derive the ψ weights given the π weights and vice versa (Box & Jenkins 1976:48).

7.1. Stationary Conditions. We have from Theorem 1.13 that the general linear model will be stationary if the ψ_i weights are square summable. Box and Jenkins (1976:49), following a slightly different approach, showed that the model will be stationary under the following conditions.

THEOREM 1.61. *The general linear model will be stationary if the series* (1.87)

$$\psi(z)$$

converge on or within the unit circle

$$|z| \leq 1.$$

PROOF. See Box & Jenkins (1976:80-82). □

7.2. Invertibility Conditions. In order to be able to invert the system we require certain restrictions on the π_i weights. These restrictions are independent to the stationarity conditions and also apply to e.g. non stationary time series.

We have seen in Section 6.3 that we require the eigenvalues of the $\mathbf{F}$ matrix to lie within the unit circle in order to be able to express the then stable difference equation as an infinite sum of the ε_t . This is exactly the invertibility conditions and we thus require from Corollary 1.59 that the roots of the characteristic equation (1.85) lie outside the unit circle in order to invert the system.

Box and Jenkins (1976:51) formulated the invertibility conditions slightly differently and required that $\pi(z)$ converges on or within the unit circle. Using their approach we require that $\psi(z)$ converge on or within the unit circle for stationarity and that $\pi(z)$ converge on or within the unit circle for invertibility.

8. ARMA Models

The models used by Box and Jenkins (1976) to describe homogeneous non stationary time series behaviour belong to the so called ARIMA or (*Autoregressive Integrated Moving Average*) class of models. If only stationary time series are considered, or if it is assumed that the time series under consideration has been transformed through differencing or some suitable method to a stationary time series, then only ARMA or *Autoregressive Moving Average* models are needed. We can further subdivide ARMA models into AR or *autoregressive* and MA or *moving average* models.

8.1. AR Models. We will first of all consider autoregressive or AR models.

DEFINITION 1.62. If a time series⁵ y_t is linearly dependent on p of its previous observations

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \quad (1.90)$$

with ε_t white noise it belongs to the AR(p) or autoregressive class of models of order p .

It follows that AR(p) models have a lot in common with difference equations such as (1.54).

8.1.1. *The Markov Model.* The simplest of all AR(p) models is the AR(1) or Markov model.

DEFINITION 1.63. The AR(1) model is defined as:

$$y_t = \delta + \phi_1 y_{t-1} + \varepsilon_t. \quad (1.91)$$

This is similar to the first order difference equation and from Theorem 1.44 we have the stationary condition:

THEOREM 1.64. An AR(1) model is stationary if, and only if,

$$|\phi_1| < 1. \quad (1.92)$$

⁵The notation followed by Box & Jenkins (1976) and many others is to write z_t instead of y_t and a_t instead of ε_t .

THEOREM 1.65. *The mean of a stationary AR(1) model (1.91) is*

$$\mu = \frac{\delta}{1 - \phi_1}.$$

PROOF. Since y_t is stationary we have

$$E[y_t] = E[\delta + \phi_1 y_{t-1} + \varepsilon_t]$$

$$\text{thus} \quad \mu = \delta + \phi_1 \mu$$

$$\text{or} \quad \mu = \frac{\delta}{1 - \phi_1}.$$

□

COROLLARY 1.66. *We may use this mean to simplify the model through*

$$z_t = y_t - \mu \tag{1.93}$$

yielding the zero mean AR(1) model

$$z_t = \phi_1 z_{t-1} + \varepsilon_t.$$

PROOF. Substituting (1.93) into (1.91) we find

$$y_t - \mu = \delta + \phi_1 y_{t-1} + \varepsilon_t - \mu$$

$$\begin{aligned} \text{or} \quad z_t &= \delta + \phi_1(y_{t-1} - \mu) + (\phi_1 - 1)\mu + \varepsilon_t \\ &= \delta + \phi_1 z_{t-1} - \delta + \varepsilon_t \\ &= \phi_1 z_{t-1} + \varepsilon_t. \end{aligned}$$

□

This transformation has the effect of setting the mean to zero and greatly simplifies further developments without loss of generality. As usual we shall indicate the actual time series with y_t and the transformed, zero mean time series with z_t .

THEOREM 1.67. *The variance of the AR(1) model is*

$$\gamma_0 = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2}. \tag{1.94}$$

PROOF.

$$\begin{aligned}
 \gamma_0 &= E[z_t^2] \\
 &= E[(\phi_1 z_{t-1} + \varepsilon_t)^2] \\
 &= \phi_1^2 E[z_{t-1}^2] + 2\phi_1 E[z_{t-1}\varepsilon_t] + E[\varepsilon_t^2] \\
 &= \phi_1^2 \gamma_0 + \sigma_\varepsilon^2 \\
 \text{thus } \gamma_0 &= \frac{\sigma_\varepsilon^2}{1 - \phi_1^2}.
 \end{aligned}$$

□

COROLLARY 1.68. *Since the variance of a stationary model is both finite and positive, we must have $|\phi_1| < 1$.*

This is again the stationary condition of an AR(1) model.

THEOREM 1.69. *The autocovariance of a stationary AR(1) model is*

$$\gamma_k = \phi_1^k \gamma_0. \quad (1.95)$$

PROOF. The first order autocovariance is

$$\begin{aligned}
 \gamma_1 &= E[z_t z_{t-1}] \\
 &= E[(\phi_1 z_{t-1} + \varepsilon_t) z_{t-1}] \\
 &= \phi_1 E[z_{t-1} z_{t-1}] + E[\varepsilon_t z_{t-1}] \\
 &= \phi_1 \gamma_0.
 \end{aligned}$$

Similarly we have

$$\begin{aligned}
 \gamma_k &= \phi_1 \gamma_{k-1} \\
 &= \phi_1 \phi_1 \gamma_{k-2} \\
 &\vdots \\
 &= \phi_1^k \gamma_0.
 \end{aligned}$$

□

COROLLARY 1.70. As $\rho_k = \frac{\gamma_k}{\gamma_0}$ we may use (1.95) to obtain the autocorrelation of a stationary AR(1) model as

$$\rho_k = \phi_1^k. \quad (1.96)$$

THEOREM 1.71. The partial autocorrelation of the AR(1) model is

$$\phi_{kk} = \begin{cases} \rho_1 & k = 1 \\ 0 & k > 1 \end{cases}. \quad (1.97)$$

PROOF. The first order partial autocorrelation is calculated using (1.21) as simply $\phi_{11} = \rho_1$. For the second and third partial autocorrelations we make use of (1.22) and (1.23). Since

$$\rho_2 = \rho_1 \rho_1, \quad \rho_3 = \rho_1 \rho_2, \quad \dots$$

the determinant in the numerator of (1.22) and (1.23) becomes zero as the last column is ρ_1 times the first. The same is true for further partial autocorrelations so that $\phi_{kk} = 0$ for the AR(1) model if $k > 1$. $\square$

8.1.2. *The Random Walk Model.* The random walk – mentioned earlier as a first order difference equation with unit root – is just a simplified AR(1) model.

DEFINITION 1.72. The *random walk* is defined as

$$y_t = \delta + y_{t-1} + \varepsilon_t. \quad (1.98)$$

If $\delta = 0$ we obtain a random walk with no drift, and if $\delta \neq 0$ we have a random walk with drift. Examples of such models are shown in Figures 1.8 and 1.9 respectively.

The random walk is a neatly balanced situation between stationary and non stationary AR(1) models. If the time series starts at $t = 0$ then

$$\begin{aligned} y_t &= \delta + y_{t-1} + \varepsilon_t \\ &= \delta + \delta + y_{t-2} + \varepsilon_t + \varepsilon_{t-1} \\ &\vdots \\ &= y_0 + t\delta + \sum_{i=0}^{t-1} \varepsilon_{t-i}. \end{aligned} \quad (1.99)$$

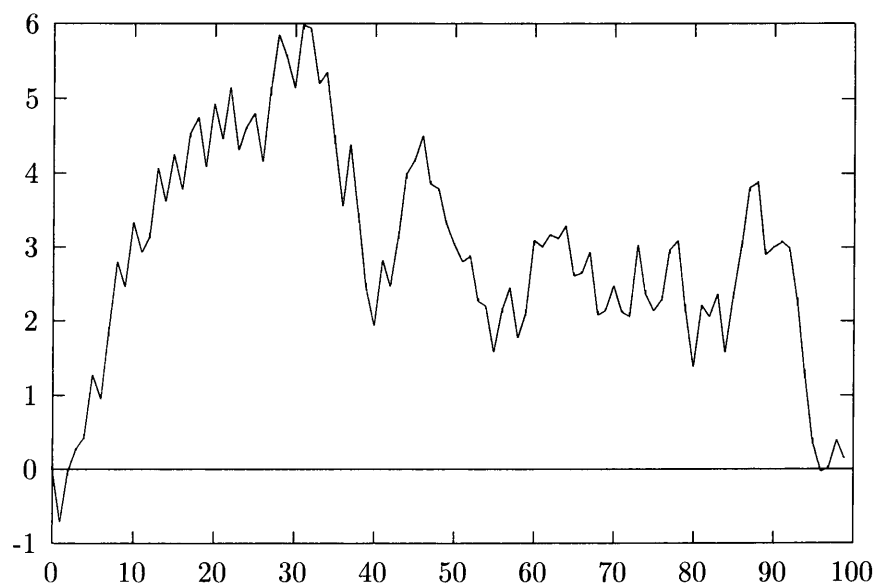


FIGURE 1.8. Random Walk with No Drift

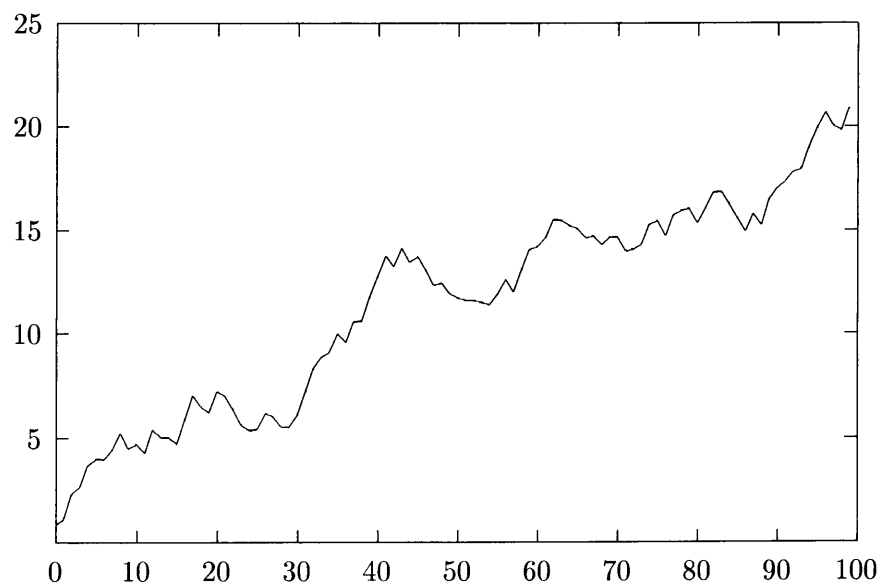


FIGURE 1.9. Random Walk with Drift

THEOREM 1.73. *The mean of the random walk at time t is*

$$\mu_t = y_0 + t\delta. \quad (1.100)$$

PROOF. The result is obtained by substituting $E[\varepsilon_t] = 0$ into (1.99). $\square$

THEOREM 1.74. *The variance of the random walk at time t is*

$$\gamma_0 = t\sigma_\varepsilon^2. \quad (1.101)$$

PROOF. We have

$$\begin{aligned} \gamma_0 &= E[y_t^2] - E^2[y_t] \\ &= E \left[\left(y_0 + t\delta + \sum_{i=0}^{t-1} \varepsilon_{t-i} \right)^2 \right] - (y_0 + t\delta)^2 \\ &= E \left[\left(\sum_{i=0}^{t-1} \varepsilon_{t-i} \right)^2 \right] \\ &= t\sigma_\varepsilon^2. \end{aligned}$$

□

This confirms the non stationarity of the random walk.

THEOREM 1.75. *The autocovariance of the random walk is*

$$\gamma_k = (t - k)\sigma_\varepsilon^2. \quad (1.102)$$

PROOF. The first order autocovariance is

$$\begin{aligned} \gamma_1 &= E[y_t y_{t-1}] - E[y_t] E[y_{t-1}] \\ &= E \left[\left(y_0 + t\delta + \sum_{i=0}^{t-1} \varepsilon_{t-i} \right) \left(y_0 + (t-1)\delta + \sum_{i=1}^{t-1} \varepsilon_{t-i} \right) \right] \\ &\quad - (y_0 + t\delta)(y_0 + (t-1)\delta) \\ &= E \left[\left(\sum_{i=0}^{t-1} \varepsilon_{t-i} \right) \left(\sum_{i=1}^{t-1} \varepsilon_{t-i} \right) \right] + (y_0 + t\delta)(y_0 + (t-1)\delta) \\ &\quad + (y_0 + t\delta) E \left[\sum_{i=1}^{t-1} \varepsilon_{t-i} \right] + (y_0 + (t-1)\delta) E \left[\sum_{i=0}^{t-1} \varepsilon_{t-i} \right] \\ &\quad - (y_0 + t\delta)(y_0 + (t-1)\delta) \\ &= E \left[\left(\sum_{i=0}^{t-1} \varepsilon_{t-i} \right) \left(\sum_{i=1}^{t-1} \varepsilon_{t-i} \right) \right] \\ &= E[(\varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_1)(\varepsilon_{t-1} + \varepsilon_{t-2} + \cdots + \varepsilon_1)] \\ &= (t-1)\sigma_\varepsilon^2. \end{aligned}$$

Similarly we have

$$\begin{aligned}\gamma_k &= E[(\varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_1)(\varepsilon_{t-k} + \varepsilon_{t-2} + \cdots + \varepsilon_1)] \\ &= (t-k)\sigma_\varepsilon^2.\end{aligned}$$

□

COROLLARY 1.76. *The autocorrelation of the random walk is*

$$\rho_k = \frac{t-k}{t}. \quad (1.103)$$

If t is sufficiently large we have a situation where all autocorrelations of the random walk tend to unity. Thus the random walk is exceptionally smooth, even though it is non stationary.

Also note that (1.98) tells us that the *best guess* that we can make for y_{t+1} at time t is simply the conditional expectation which will be y_t for a random walk with no drift – thus it is a *martingale*. The random walk is a natural starting point for modeling stock prices (with drift) or changes in stock prices (no drift) under the assumption of *efficient markets* – a situation that arises if all available information at time t is already reflected in the stock price y_t and any future changes are unpredictable and the result of random shocks.

8.1.3. *The effects of aggregation.* Surprisingly, if a stock price time series may be modeled using a random walk then there is an argument for using daily (or weekly or monthly etc.) *closes* rather than *averages* when quoting prices as is also general market practice. We can write the $m > 1$ successive terms in the random walk *with no drift* as (Kendall & Ord 1973:58)

$$y_t, \quad y_t + \varepsilon_{t+1}, \quad y_t + \varepsilon_{t+1} + \varepsilon_{t+2}, \quad \dots, \quad y_t + \varepsilon_{t+1} + \varepsilon_{t+2} + \cdots + \varepsilon_{t+m-1}.$$

The next m successive terms will then be

$$y_t + \varepsilon_{t+1} + \varepsilon_{t+2} + \cdots + \varepsilon_{t+m}, \quad \dots, \quad y_t + \varepsilon_{t+1} + \varepsilon_{t+2} + \cdots + \varepsilon_{t+2m-1}.$$

If we define the averages of these sets of successive values as

$$\bar{y}_{(1)} = \sum_{i=0}^{m-1} \frac{y_{t+i}}{m}$$

and, similarly

$$\bar{y}_{(2)} = \sum_{i=0}^{m-1} \frac{y_{t+m+i}}{m}$$

or, in general

$$\bar{y}_{(k)} = \sum_{i=0}^{m-1} \frac{y_{t+(k-1)m+i}}{m}$$

and also define

$$\begin{aligned} d_{(1)} &= \bar{y}_{(2)} - \bar{y}_{(1)} \\ &= \frac{1}{m} (\varepsilon_{t+1} + 2\varepsilon_{t+2} + \cdots + (m-1)\varepsilon_{t+m-1} + m\varepsilon_{t+m} \\ &\quad + (m-1)\varepsilon_{t+m+1} + \cdots + 2\varepsilon_{t+2m-2} + \varepsilon_{t+2m-1}) \end{aligned}$$

so that we have in general

$$\begin{aligned} d_{(k)} &= \bar{y}_{(k+1)} - \bar{y}_{(k)} \\ &= \frac{1}{m} (\varepsilon_{t+(k-1)m+1} + 2\varepsilon_{t+(k-1)m+2} + \cdots + (m-1)\varepsilon_{t+(k-1)m+m-1} \\ &\quad + m\varepsilon_{t+km} + (m-1)\varepsilon_{t+km+1} + \cdots + 2\varepsilon_{t+(k+1)m-2} + \varepsilon_{t+(k+1)m-1}). \end{aligned}$$

If we take say $m = 4$ we can construct Table 1.7. Thus if we want the coefficient of e.g. ε_{t+5} in $d_{(2)}$ we see from this table that it is $\frac{1}{4}$. Using these definitions and Table 1.7 as a guide we show the following.

THEOREM 1.77. *The mean of $d_{(k)}$ is zero.*

PROOF. It is clear from the definition of $d_{(k)}$ as can be verified from Table 1.7 that it is a weighted sum of ε_t . Since $E[\varepsilon_t] = 0$ the result follows. $\square$

THEOREM 1.78. *The variance of $d_{(k)}$ is*

$$\text{Var}(d_{(k)}) = \sigma_\varepsilon^2 \cdot \frac{2m^2 + 1}{3m}. \quad (1.104)$$

Variable	y_t	ε_{t+1}	ε_{t+2}	ε_{t+3}	ε_{t+4}	ε_{t+5}	ε_{t+6}	ε_{t+7}	ε_{t+8}	ε_{t+9}	ε_{t+10}	
y_t	1											
y_{t+1}	1	1										
y_{t+2}	1	1	1									
y_{t+3}	1	1	1	1								
$\bar{y}_{(1)}$	$\frac{4}{4}$	$\frac{3}{4}$	$\frac{2}{4}$	$\frac{1}{4}$								
y_{t+4}	1	1	1	1	1							
y_{t+5}	1	1	1	1	1	1						
y_{t+6}	1	1	1	1	1	1	1					
y_{t+7}	1	1	1	1	1	1	1	1				
$\bar{y}_{(2)}$	$\frac{4}{4}$	$\frac{4}{4}$	$\frac{4}{4}$	$\frac{4}{4}$	$\frac{4}{4}$	$\frac{3}{4}$	$\frac{2}{4}$	$\frac{1}{4}$				
$d_{(1)}$	0	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{3}{4}$	$\frac{4}{4}$	$\frac{3}{4}$	$\frac{2}{4}$	$\frac{1}{4}$				
y_{t+8}	1	1	1	1	1	1	1	1	1			
y_{t+9}	1	1	1	1	1	1	1	1	1	1		
y_{t+10}	1	1	1	1	1	1	1	1	1	1	1	
y_{t+11}	1	1	1	1	1	1	1	1	1	1	1	
$\bar{y}_{(3)}$	$\frac{4}{4}$	$\frac{4}{4}$	$\frac{4}{4}$	$\frac{4}{4}$	$\frac{4}{4}$	$\frac{4}{4}$	$\frac{4}{4}$	$\frac{4}{4}$	$\frac{4}{4}$	$\frac{3}{4}$	$\frac{2}{4}$	$\frac{1}{4}$
$d_{(2)}$	0	0	0	0	0	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{3}{4}$	$\frac{4}{4}$	$\frac{3}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

TABLE 1.7. Sample Coefficients of y , $\bar{y}$ and d with $m = 4$

PROOF. In order to calculate the variance we make use of

$$\text{Cov}(\varepsilon_t, \varepsilon_{t-k}) = \begin{cases} \sigma_\varepsilon^2, & k = 0 \\ 0, & k \neq 0 \end{cases}.$$

When calculating the variance we only need the squares of the coefficients of the i.i.d. ε_t that make up $d_{(k)}$.

$$\begin{aligned}\text{Var}(d_{(k)}) &= E[d_{(k)}^2] \\ &= \frac{\sigma_\varepsilon^2}{m^2} (1^2 + 2^2 + \cdots + (m-1)^2 + m^2 \\ &\quad + (m-1)^2 + \cdots + 2^2 + 1^2) \\ &= \frac{\sigma_\varepsilon^2}{m^2} (m^2 + 2 \cdot \sum_{i=1}^{m-1} i^2) \\ &= \sigma_\varepsilon^2 \cdot \frac{2m^2 + 1}{3m}\end{aligned}$$

since

$$\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}. \quad (1.105)$$

□

LEMMA 1.79.

$$\sum_{i=1}^n i \cdot (n-i) = \frac{n^3 - n}{6}. \quad (1.106)$$

PROOF.

$$\begin{aligned}\sum_{i=1}^n i \cdot (n-i) &= n \sum_{i=1}^n i - \sum_{i=1}^n i^2 \\ &= \frac{n^2(n+1)}{2} - \frac{n(n+1)(2n+1)}{6} \\ &= \frac{n^3 - n}{6}.\end{aligned}$$

□

THEOREM 1.80. *The first order autocovariance of $d_{(k)}$ is*

$$\text{Cov}(d_{(k)}, d_{(k-1)}) = \sigma_\varepsilon^2 \cdot \frac{m^2 - 3m + 2}{6m}. \quad (1.107)$$

PROOF. Similarly to the previous proof we combine the coefficients of the same ε_t that we have in the $d_{(k)}$ spaced one period apart. This can be verified from Table 1.7 and using $d_{(1)}$ and $d_{(2)}$. The only ε_t common to both

these $d_{(k)}$ are $\varepsilon_{t+5}, \varepsilon_{t+6}$ and ε_{t+7} and only their coefficients are used. In general we have

$$\begin{aligned}\text{Cov}(d_{(k)}, d_{(k-1)}) &= \sigma_\varepsilon^2 \cdot \frac{1(m-1) + 2(m-1) + \cdots + (m-1)1}{m^2} \\ &= \sigma_\varepsilon^2 \cdot \frac{\sum_{i=1}^{m-1} i \cdot (m-i)}{m^2}\end{aligned}$$

and, making use of the identity (1.106) this becomes

$$= \sigma_\varepsilon^2 \cdot \frac{m^2 - 3m + 2}{6m}.$$

□

THEOREM 1.81. *Also note that similarly higher autocovariances are zero*

$$\text{Cov}(d_{(k)}, d_{(k-p)}) = 0, \quad k > p. \quad (1.108)$$

COROLLARY 1.82. *The autocorrelation of $d_{(k)}$ of order p is*

$$\text{Cor}(d_{(k)}, d_{(k-p)}) = \begin{cases} 1, & p = 0 \\ \frac{1}{2} \cdot \frac{m^2 - 3m + 2}{2m^2 + 1}, & p = 1 \\ 0, & p > 1 \end{cases} \quad (1.109)$$

Although first differences of the original time series

$$y_t - y_{t-1} = \delta + \varepsilon_t$$

are uncorrelated those of the aggregated series $d_{(k)}$ are not. If we apply l'Hôpital's rule twice we can calculate the limit of (1.109) for $p = 1$ as

$$\rho_1 = \lim_{m \rightarrow \infty} \frac{1}{2} \cdot \frac{m^2 - 3m + 2}{2m^2 + 1} = \frac{1}{4}$$

and thus as m or the number of stock prices recorded per day increases the first order autocorrelation⁶ of the average of these observations will approach $\frac{1}{4}$. This *effect of aggregation* is one reason why stock prices are quoted as closes rather than averages.

⁶Kendall and Ord (1990:58) calculated the first order autocovariance of $d_{(k)}$ as

$$\sigma_\varepsilon^2 \cdot \frac{m^2 - 1}{6m}$$

through (erroneously?) including an additional m^2 in the sum of the coefficients in (1.107). Even if this is used the limit of the first order correlation of the $d_{(k)}$ still is $\frac{1}{4}$.

8.1.4. *The Yule Model.* If y_t depends on a linear combination of two previous observations it is called an AR(2) or Yule model.

DEFINITION 1.83. The AR(2) or Yule model is defined as

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t. \quad (1.110)$$

THEOREM 1.84. *The AR(2) model is stationary if, and only if, the roots of the characteristic equation*

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 \quad (1.111)$$

are outside the unit circle.

PROOF. The eigenvalues of the matrix $\mathbf{F}$ are within the unit circle since the roots of the characteristic equation are outside by Corollary 1.60. Then from Theorem 1.56 the inverse $(\mathcal{I} - \mathbf{F})^{-1}$ exists with $(1, 1)$ element

$$\frac{1}{1 - \phi_1 - \phi_2}.$$

In order to show stationarity we write the system as an infinite sum of past errors with weights ψ_i . From (1.57) ψ_i is the $(1, 1)$ element of the matrix $\mathbf{F}^i$. We write

$$\begin{aligned} \psi(B) &= \phi^{-1}(B) \\ &= \frac{1}{1 - \phi_1 B - \phi_2 B^2} \\ &= \psi_0 + \psi_1 B + \psi_2 B^2 + \dots \end{aligned}$$

and multiply (1.110) with this to obtain

$$y_t \phi(B) \psi(B) = (\delta + \varepsilon_t) \psi(B)$$

or

$$y_t = \delta \psi(B) + \varepsilon_t \psi(B).$$

It now follows that

$$\delta \psi(B) = \frac{\delta}{1 - \phi_1 - \phi_2}$$

and we only have to show that the infinite absolute sum

$$\sum_{i=0}^{\infty} |\psi_i|$$

is finite. Stationarity then follows by Corollary 1.14.

To show that the absolute sum converges we rewrite the characteristic equation by Theorem 1.58 as

$$1 - \phi_1 z - \phi_2 z^2 = (1 - \lambda_1 z)(1 - \lambda_2 z)$$

with λ_i inside the unit circle only if ϕ_i are outside. These two λ_i are the roots of a quadratic equation and either distinct real numbers, complex conjugates or the same real number (de la Rosa et al. 1987:90).

1. If λ_1 and λ_2 are real and distinct we have from (1.61)

$$\begin{aligned} \sum_{i=0}^{\infty} |\psi_i| &= \sum_{i=0}^{\infty} |c_1 \lambda_1^i + c_2 \lambda_2^i| \\ &\leq \sum_{i=0}^{\infty} |c_1 \lambda_1^i| + \sum_{i=0}^{\infty} |c_2 \lambda_2^i| \\ &= \frac{|c_1|}{1 - |\lambda_1|} + \frac{|c_2|}{1 - |\lambda_2|} \end{aligned}$$

which is finite if, and only if, $|\lambda_i| < 1$.

2. If λ_1 and λ_2 are complex conjugates it follows from Corollary 1.53 that c_1 and c_2 are also complex conjugates, say

$$\begin{aligned} c_1 &= \alpha + \iota\beta, \quad \text{with } \iota = \sqrt{-1} \\ c_2 &= \alpha - \iota\beta. \end{aligned}$$

If we write a complex number in modulus-argument form and raise it to an integer power after de Moivre's theorem (de la Rosa et al. 1987:63)

$$\lambda_i^k = d(\cos(\theta k) + \iota \sin(\theta k))$$

$$\text{with } d = |\lambda_1| = |\lambda_2| < 1$$

we have, using k as index variable to ease reading and $\iota = \sqrt{-1}$

$$\begin{aligned}
\sum_{k=0}^{\infty} |\psi_k| &= \sum_{k=0}^{\infty} |c_1 \lambda_1^k + c_2 \lambda_2^k| \\
&= \sum_{k=0}^{\infty} d^k \left| [(\alpha + \iota\beta)(\cos(k\theta) + \iota \sin(k\theta)) \right. \\
&\quad \left. + (\alpha - \iota\beta)(\cos(k\theta) - \iota \sin(k\theta))] \right| \\
&= \sum_{k=0}^{\infty} d^k \left| [\alpha \cos(k\theta) + \iota\alpha \sin(k\theta) + \iota\beta \cos(k\theta) - \beta \sin(k\theta) \right. \\
&\quad \left. + \alpha \cos(k\theta) - \iota\alpha \sin(k\theta) - \iota\beta \cos(k\theta) - \beta \sin(k\theta)] \right| \\
&= \sum_{k=0}^{\infty} d^k \left| [2\alpha \cos(k\theta) - 2\beta \sin(k\theta)] \right| \\
&\leq |2\alpha| \sum_{k=0}^{\infty} d^k |\cos(k\theta)| + |2\beta| \sum_{k=0}^{\infty} d^k |\sin(k\theta)| \\
&\leq |2\alpha| \sum_{k=0}^{\infty} d^k + |2\beta| \sum_{k=0}^{\infty} d^k \\
&= 2(|\alpha| + |\beta|) \sum_{k=0}^{\infty} d^k = 2 \cdot \frac{|\alpha| + |\beta|}{1-d}
\end{aligned}$$

which is finite if, and only if, $|\lambda_i| < 1$.

3. Finally we have the possibility of a repeated real eigenvalue $\lambda_1 = \lambda_2 = \lambda$. Then from (1.74) we calculate (Hamilton 1994:770)

$$\begin{aligned}
\sum_{k=0}^{\infty} |\psi_k| &= \sum_{k=0}^{\infty} |c_1 \lambda^k + c_2 k \lambda^{k-1}| \leq |c_1| \sum_{k=0}^{\infty} |\lambda^k| + |c_2| \sum_{k=0}^{\infty} |k \lambda^{k-1}| \\
&= \frac{|c_1|}{1-|\lambda|} + |c_2| \cdot \left[1 + (|\lambda| + |\lambda|) + (|\lambda|^2 + |\lambda|^2 + |\lambda|^2) \right. \\
&\quad \left. + (|\lambda|^3 + |\lambda|^3 + |\lambda|^3 + |\lambda|^3) + \dots \right] \\
&= K_1 + |c_2| \cdot \left[(1 + |\lambda| + |\lambda|^2 + |\lambda|^3 + \dots) \right. \\
&\quad \left. + (|\lambda| + |\lambda|^2 + |\lambda|^3 + |\lambda|^4 + \dots) + \dots \right] \\
&= K_1 + |c_2| \cdot \left[\frac{1}{1-|\lambda|} + \frac{|\lambda|}{1-|\lambda|} + \frac{|\lambda|^2}{1-|\lambda|} + \frac{|\lambda|^3}{1-|\lambda|} + \dots \right] \\
&= K_1 + |c_2| \cdot \left[\frac{1}{1-|\lambda|} \right]^2
\end{aligned}$$

which also is finite if, and only if, $|\lambda_i| < 1$.

□

THEOREM 1.85. *The mean of a stationary AR(2) model is*

$$\mu = \frac{\delta}{1 - \phi_1 - \phi_2}. \quad (1.112)$$

PROOF.

$$E[y_t] = E[\delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t]$$

$$\text{or} \quad \mu = \delta + \phi_1 \mu + \phi_2 \mu$$

$$\text{thus} \quad \mu = \frac{\delta}{1 - \phi_1 - \phi_2}.$$

□

COROLLARY 1.86. *From the equation of the mean it follows that for the AR(2) model*

$$\delta = \mu \cdot (1 - \phi_1 - \phi_2). \quad (1.113)$$

THEOREM 1.87. *The autocovariance of a stationary AR(2) model is*

$$\gamma_k = \begin{cases} \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma_\varepsilon^2, & k = 0 \\ \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}, & k > 0 \end{cases}. \quad (1.114)$$

PROOF. If we substitute (1.113) in (1.110) we have

$$\begin{aligned} y_t &= \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \\ &= \mu \cdot (1 - \phi_1 - \phi_2) + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \end{aligned}$$

$$\text{thus} \quad y_t = \mu + \phi_1 (y_{t-1} - \mu) + \phi_2 (y_{t-2} - \mu) + \varepsilon_t$$

$$\text{or} \quad y_t - \mu = \phi_1 (y_{t-1} - \mu) + \phi_2 (y_{t-2} - \mu) + \varepsilon_t. \quad (1.115)$$

Multiplying both sides with $(y_{t-k} - \mu)$ and taking expectations yields

$$\begin{aligned} E[(y_t - \mu)(y_{t-k} - \mu)] &= E[\phi_1(y_{t-1} - \mu)(y_{t-k} - \mu) \\ &\quad + \phi_2(y_{t-2} - \mu)(y_{t-k} - \mu) + \varepsilon_t(y_{t-k} - \mu)] \\ \text{or } \gamma_k &= \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + E\left[\varepsilon_t(\phi_1(y_{t-k-1} - \mu) \right. \\ &\quad \left. + \phi_2(y_{t-k-2} - \mu) + \varepsilon_{t-k})\right] \quad \text{from (1.115)} \\ &= \begin{cases} \phi_1\gamma_1 + \phi_2\gamma_2 + \sigma_\varepsilon^2, & k = 0 \\ \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2}, & k > 0 \end{cases}. \end{aligned}$$

□

COROLLARY 1.88. *The autocorrelation of a stationary AR(2) model is*

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}. \quad (1.116)$$

COROLLARY 1.89. *The first order autocorrelation of a stationary AR(2) model is*

$$\rho_1 = \frac{\phi_1}{1 - \phi_2}. \quad (1.117)$$

PROOF.

$$\begin{aligned} \rho_1 &= \phi_1\rho_0 + \phi_2\rho_{-1} \\ &= \phi_1 + \phi_2\rho_1 \\ \text{thus } \rho_1 &= \frac{\phi_1}{1 - \phi_2}. \end{aligned}$$

□

COROLLARY 1.90. *If we substitute $\rho_0 = 1$ and (1.117) in (1.116) the second order autocorrelation of a stationary AR(2) model is*

$$\rho_2 = \frac{\phi_1^2}{1 - \phi_2} + \phi_2. \quad (1.118)$$

COROLLARY 1.91. *The variance of a stationary AR(2) model is*

$$\gamma_0 = \sigma_\varepsilon^2 \cdot \frac{1 - \phi_2}{(1 - \phi_1 - \phi_2)(1 + \phi_1 - \phi_2)(1 + \phi_2)}. \quad (1.119)$$

PROOF. We have from (1.114) and the fact that $\gamma_k = \gamma_{-k}$

$$\begin{aligned}\gamma_0 &= \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma_\varepsilon^2 \\ &= \phi_1 \gamma_0 \rho_1 + \phi_2 \gamma_0 \rho_2 + \sigma_\varepsilon^2\end{aligned}$$

and from (1.117) and (1.118)

$$= \phi_1 \gamma_0 \frac{\phi_1}{1 - \phi_2} + \phi_2 \gamma_0 \left[\frac{\phi_1^2}{1 - \phi_2} + \phi_2 \right] + \sigma_\varepsilon^2$$

thus

$$\begin{aligned}\gamma_0 &= \sigma_\varepsilon^2 \cdot \frac{1}{1 - \frac{\phi_1^2}{1 - \phi_2} - \frac{\phi_1^2 \phi_2}{1 - \phi_2} - \phi_2^2} \\ &= \sigma_\varepsilon^2 \cdot \frac{1 - \phi_2}{1 - \phi_2 - \phi_1^2 - \phi_1^2 \phi_2 - \phi_2^2 + \phi_2^3} \\ &= \sigma_\varepsilon^2 \cdot \frac{1 - \phi_2}{(1 - \phi_1 - \phi_2)(1 + \phi_1 - \phi_2)(1 + \phi_2)}.\end{aligned}$$

□

THEOREM 1.92. *The partial autocorrelation of a stationary AR(2) model is*

$$\phi_{kk} = \begin{cases} \rho_1 & k = 1 \\ \phi_2 & k = 2 \\ 0 & k \geq 3 \end{cases} \quad (1.120)$$

PROOF. For $k = 1$ the result follows immediately from (1.21) while, as with the AR(1) model and following from (1.116), we have a numerator of zero in (1.23) and higher $k \geq 3$ autocorrelations so that $\phi_{kk} = 0$, $k \geq 3$.

For $k = 2$ we substitute (1.117) and (1.118) into (1.22)

$$\begin{aligned}
 \phi_{22} &= \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \\
 &= \frac{\phi_2 + \frac{\phi_1^2}{1-\phi_2} - \left[\frac{\phi_1}{1-\phi_2}\right]^2}{1 - \left[\frac{\phi_1}{1-\phi_2}\right]^2} \\
 &= \frac{\phi_2(1-\phi_2)^2 + \phi_1^2(1-\phi_2) - \phi_1^2}{(1-\phi_2)^2 - \phi_1^2} \\
 &= \phi_2.
 \end{aligned}$$

□

The partial autocorrelation is also a correlation and less than one in magnitude. Thus $|\phi_{kk}| < 1$ which implies

$$\begin{aligned}
 |\phi_2| &< 1, \\
 \phi_2 + \phi_1 &< 1
 \end{aligned}$$

and

$$\phi_2 - \phi_1 < 1.$$

These restrictions define a triangular range of allowable values for ϕ_1 and ϕ_2 as depicted in Figure 1.10 and can be used in practice to quickly verify stationarity.

8.1.5. *General AR(p) Models.* We defined the AR(p) model in (1.90) and now derive some of its properties.

THEOREM 1.93. *The AR(p) model is stationary if, and only if, the roots of the characteristic equation*

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p \quad (1.121)$$

lie outside the unit circle.

PROOF. The result follows along the same lines of that of Theorem 1.84. The eigenvalues λ_i are inside the unit circle only if the roots of the characteristic equation are outside and as before we can write the AR(p) model

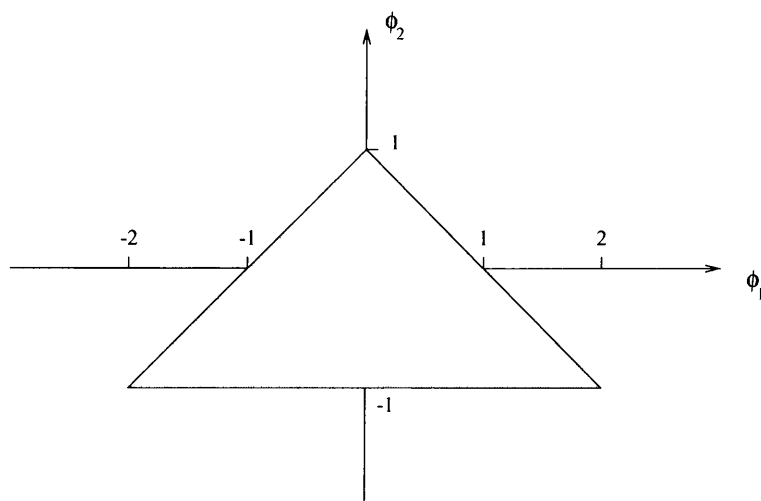


FIGURE 1.10. Parameter Range of Stationary AR(2) Model

as

$$\begin{aligned}
 y_t &= \psi(B)\delta + \psi(B)\varepsilon_t \\
 &= \frac{\delta}{1 - \phi_1 - \phi_2 - \dots - \phi_p} + \sum_{i=0}^{\infty} \psi_i^i \varepsilon_{t-i}.
 \end{aligned}$$

For stationarity we need the sum of the absolute of the ψ_i weights to converge. This is easy to show as in Theorem 1.84 if the eigenvalues are real and distinct or complex conjugates or any combination thereof. If $q \leq p$ of the eigenvalues are repeated and the rest $p - q$ of the eigenvalues is some combination of distinct real numbers and complex conjugates then it is not difficult to show, again along the lines of Theorem 1.84 that the portion of the absolute sum of the ψ_i weights attributable to this combination of distinct eigenvalues is finite.

We need to show that the portion that each of the repeated eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$ with $k < q$ contributes to the absolute sum is finite. We only need to show that this contribution is finite for one of the repeated eigenvalues where this eigenvalue is repeated an arbitrary number of times, say k . Then the contribution of any of the other repeated eigenvalues will also be finite so that the total contribution due to repeated eigenvalues or equivalently the sum of these individual contributions will be finite.

If the real eigenvalue λ is repeated k times then from Theorem 1.54 its contribution to the absolute sum is not more than

$$K = \sum_{i=0}^{\infty} \sum_{j=0}^{k-1} |c_j \binom{i}{j} \lambda^{i-j}|.$$

If we define

$$K_l = \sum_{i=0}^{\infty} \sum_{j=0}^{l-1} |c_j \binom{i}{j} \lambda^{i-j}|, \quad l = 2, 3, \dots, k$$

as the contribution of l of the repeated eigenvalue λ where $l \leq k$ we need to show that K_k is finite. We have already shown that K_2 is finite in Theorem 1.84. If we assume that K_l is finite we now have to show that K_{l+1} is finite.

$$\begin{aligned} K_{l+1} &= \sum_{i=0}^{\infty} \sum_{j=0}^l |c_j \binom{i}{j} \lambda^{i-j}| \\ &= \sum_{i=0}^{\infty} \left[\sum_{j=0}^{l-1} |c_j \binom{i}{j} \lambda^{i-j}| + |c_l \binom{i}{l} \lambda^{i-l}| \right] \\ &= K_l + \sum_{i=0}^{\infty} |c_l \binom{i}{l} \lambda^{i-1}|. \end{aligned}$$

Since K_l is finite by assumption we only have to show that

$$\sum_{i=0}^{\infty} |a_i| = \sum_{i=0}^{\infty} |c_l \binom{i}{l} \lambda^{i-l}| \quad (1.122)$$

is finite. The limit of the absolute ratio between a_{i+1} and a_i is

$$\begin{aligned} \lim_{i \rightarrow \infty} \left| \frac{a_{i+1}}{a_i} \right| &= \lim_{i \rightarrow \infty} \left| \frac{c_l \binom{i+1}{l} \lambda^{i+1-l}}{c_l \binom{i}{l} \lambda^{i-l}} \right| \\ &= \lim_{i \rightarrow \infty} |\lambda| \cdot \frac{\frac{(i+1)!}{k!(i+1-k)!}}{\frac{i!}{k!(i-k)!}} \\ &= |\lambda| \lim_{i \rightarrow \infty} \frac{i+1}{i+1-k} \\ &= |\lambda| < 1. \end{aligned}$$

only if $|\lambda|$ is inside the unit circle. From the well known *Generalized Ratio Test* (Ellis & Gulick 1986:537) the series (1.122) converges absolutely if the ratio is less than 1 and the result follows. $\square$

We can therefore write for a stationary (and stable) AR(p) model

$$\phi^{-1}(B) = \psi(B)$$

$$\frac{1}{1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p} = \psi_0 + \psi_1 B + \psi_2 B^2 + \dots$$

or

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(\psi_0 + \psi_1 B + \psi_2 B^2 + \dots) = 1 \quad (1.123)$$

which we may use to recursively calculate the ψ_i weights. Since the coefficient of B^0 must be 1 and that of B^k with $k > 1$ must be zero from (1.123) we have

$$\psi_0 = 1$$

$$\psi_1 = \phi_1 \quad \text{from} \quad (-\phi_1 + \psi_1) B = 0 \cdot B$$

$$\psi_2 = \phi_1^2 + \phi_2 \quad \text{from} \quad (-\phi_2 + \psi_2 - \phi_1 \psi_1) B^2 = 0 \cdot B^2$$

and so forth.

We now calculate the mean, autocovariance and autocorrelation of a stationary AR(p) model. The proofs of the following theorems are along the lines of those for an AR(2) model and only the results are stated here.

THEOREM 1.94. *The mean of a stationary AR(p) model is*

$$\mu = \frac{\delta}{1 - \phi_1 - \phi_2 - \dots - \phi_p}. \quad (1.124)$$

THEOREM 1.95. *The variance of a stationary AR(p) model is*

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \dots + \phi_p \gamma_p + \sigma_\varepsilon^2. \quad (1.125)$$

THEOREM 1.96. *The autocovariance of a stationary AR(p) model satisfies the same difference equation as the model itself*

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p}. \quad (1.126)$$

COROLLARY 1.97. *The autocorrelation of a stationary AR(p) model also satisfies*

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p}. \quad (1.127)$$

COROLLARY 1.98. *The variance of a stationary AR(p) model can also be written as*

$$\gamma_0 = \frac{\sigma_\varepsilon^2}{1 - \phi_1\rho_1 - \phi_2\rho_2 - \cdots - \phi_p\rho_p}. \quad (1.128)$$

THEOREM 1.99. *The partial autocorrelation of a stationary AR(p) model has cut off after lag p*

$$\phi_{kk} = 0, \quad k > p. \quad (1.129)$$

8.2. MA Models. *Moving averages* are used extensively in the financial world in order to identify trends, smooth volatile data, as decision making tools in trading strategies and so on. If moving averages are thus employed one has to guard against the effects of aggregation discussed in the previous section or the emergence of the so called *Slutzky-Yule* effect (discussed later in this section). A different but related concept is that of time series moving average or MA models.

In the previous section we studied AR(p) models that depended on p of the previous values of the time series itself. It makes sense to also use the previous error terms to model the behaviour of the time series. A good example where this applies is the South African quarterly maize crop.

Depending on the weather, the bulk of this crop is produced in either the second or the third quarter of each year. If the crop is expected in the second quarter but only ripens and gets harvested towards the third quarter, our forecasting model will produce a fairly large and negative error in the second quarter. When the crop does get harvested this second quarter error will be followed by a fairly large and positive error in the third quarter. Since the errors of an accurate model should be uncorrelated white noise this makes out an argument for the inclusion of previous errors in the model itself.

As another example consider the headline inflation figures defined in Section 2.3. If the government unexpectedly increases sales tax during a given month our time series model is bound to produce a fairly large and positive error. One year down the line we will observe a sudden drop in

<i>Quarterly Consumer Price Index (CPI)</i>				
Year	Q1	Q2	Q3	Q4
1	100.00 (+5%)	105.00 (+5%)	110.25 (+5%)	115.76 (+5%)
2	121.55 (+5%)	127.63 (+5%)	140.39 (+10%)	147.41 (+5%)
3	154.78 (+5%)	162.52 (+5%)	170.65 (+5%)	179.18 (+5%)
4	188.14 (+5%)	197.54 (+5%)	207.42 (+5%)	217.79 (etc.)

<i>Quarterly Inflation Rate</i>				
(Year on Year Percentage Change in CPI)				
1				
2	$\frac{121.55}{100.00} - 1$ 21.6%	$\frac{127.63}{105.00} - 1$ 21.6%	$\frac{140.39}{110.25} - 1$ 27.3%	$\frac{147.41}{115.76} - 1$ 27.3%
3	etc. 27.3%	27.3%	21.6%	21.6%
4	21.6%	21.6%	21.6%	21.6%

TABLE 1.8. Inflation Rate in Avalon

inflation – provided that the tax rate stays the same – as the effect works itself out of the system. Table 1.8 provides an illustration of this.

Suppose that in some country, say Avalon, the quarterly inflation rate not accounting for tax increases always is 5%. Thus the consumer price index grows at a constant rate of 5% per quarter or 21.6% per year as in Table 1.8 as long as the tax rate remains unchanged. If the government decides to increase sales tax with say 5% during the third quarter of Year 2 the headline inflation during that quarter increases to 27.3%. This increase works its way into all the inflation numbers for a whole year even if we keep quarterly increases thereafter constant at 5%. This effect works itself out in due course and by the time we reach the third quarter of Year 3 the inflation figures drop back to 21.6% where it remains – until government increases taxes again.

If we model inflation or some related time series it again makes sense to include previous error terms in the model to yield better results. MA models allow us to do this.

DEFINITION 1.100. An $MA(q)$ or *Moving Average* model of order q is generated by a *weighted moving average* of q of the previous error terms

$$y_t = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \quad (1.130)$$

with ε_t white noise.

Thus an $MA(q)$ model is simply a *linear combination* of the stationary time series ε_t and we can directly employ many of the results of previous sections. In drawing on the work already done we will mostly state results as theorems and instead of providing formal proofs just refer to the supporting theorem or equation already handled. We will then investigate the *Slutsky-Yule* effect as a sample application of these models.

THEOREM 1.101. *From Corollary 1.12 it follows that a finite MA model is stationary and from Theorem 1.13 an infinite MA model is stationary if the θ_i weights are square summable.*

THEOREM 1.102. *Since $E[\varepsilon_t] = 0$ it follows that the mean of a stationary $MA(q)$ model is*

$$\mu = \delta \quad (1.131)$$

THEOREM 1.103. *If we use the mean to transform the stationary $MA(q)$ model to a zero mean model we can apply (1.7) to obtain the autocovariance as*

$$\gamma_k = \begin{cases} \sigma_\varepsilon^2(-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q), & k = 0, 1, \dots, q \\ 0, & k > q \end{cases} \quad (1.132)$$

In this expression (1.132) we set θ_0 as the negative of the coefficient of ε_{t-0} in (1.130), thus

$$\theta_0 = -1. \quad (1.133)$$

Then we can write the autocovariance (1.132) as

$$\gamma_k = \sigma_\varepsilon^2(\theta_0\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q) \quad k = 0, 1, \dots, q. \quad (1.134)$$

COROLLARY 1.104. *The variance of the stationary MA(q) model is*

$$\gamma_0 = \sigma_\varepsilon^2(1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2) \quad (1.135)$$

If we have an infinite order MA(∞) model the variance is

$$\gamma_0 = \sigma_\varepsilon^2 \lim_{q \rightarrow \infty} \left[1 + \sum_{i=1}^q \theta_i^2 \right].$$

COROLLARY 1.105. *The autocorrelation of a stationary MA(q) model is*

$$\rho_k = \begin{cases} \frac{-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2} & k = 0, 1, \dots, q \\ 0 & k > q \end{cases} \quad (1.136)$$

which cuts off at lag q .

In order to calculate the partial autocorrelations we use (1.20) or (1.24). The exact expression of the partial autocorrelation of an MA(q) processes is complicated, but it exists and is infinite in extent.

8.2.1. *The Slutsky-Yule Effect.* If we define y_t as white noise

$$y_t = \varepsilon_t$$

and form the moving average

$$\bar{y}_t = \sum_{j=0}^q l_j y_{t-j}$$

which we scale so that $l_0 = 1$ we have an MA(q) model with $\theta_j = -l_j$. The autocorrelations of this process is given by (1.136) which is not zero for all lags smaller than or equal to q .

Thus through smoothing the random series we have implicitly created autocorrelation up to lag q . Apparently systematic effects in this smoothed series may be observed due to the result of the smoothing rather than some underlying deterministic influence! This is called the *Slutsky-Yule effect* after its discoverers Slutsky (1927) and Yule (1927).

This realization that apparently systematic fluctuations can be generated as the average of random events came as a big shock to the academic world, especially since Slutsky was able to mimic an actual *trade cycle* very closely using white noise. This is the reason why many practitioners prefer

to work with data that have not been seasonally adjusted (Kendall & Ord 1990:65).

8.2.2. *Invertibility.* We now revisit the concept of invertibility. Take the first order MA(1) model with zero mean

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad (1.137)$$

which is stationary from Theorem 1.101. We can rewrite this model as an infinite order AR(∞) model provided that $|\theta_1| < 1$ since

$$y_t = (1 - \theta_1 B) \varepsilon_t$$

or

$$\begin{aligned} (1 - \theta_1 B)^{-1} y_t &= \varepsilon_t \\ (1 + \theta_1 B + \theta_1^2 B^2 + \theta_1^3 B^3 + \dots) y_t &= \varepsilon_t \end{aligned}$$

so that

$$y_t = \varepsilon_t - \theta_1 y_{t-1} - \theta_1^2 y_{t-2} - \theta_1^3 y_{t-3} - \dots \quad (1.138)$$

Here the model is invertible. If on the other hand $|\theta_1| \geq 1$ the model will be stationary but the sum (1.138) will be explosive. In order to avoid such situations we require $|\theta_1| < 1$ and refer to this as the invertibility conditions of an MA(1) model.

For a general MA(q) process we can write

$$y_t - \delta = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t \quad (1.139)$$

and then if we replace the y_t in Corollary 1.60 with the ε_t we will require that the roots of the equation

$$\theta(z) = 1 - \theta_1 z - \theta_2 z^2 - \dots - \theta_q z^q \quad (1.140)$$

lie outside the unit circle for invertibility.

8.3. ARMA Models. Fuller (1976:148-151) showed that any stationary model may be approximated accurately using either an AR or MA model of sufficiently high order. We have also discussed how to *invert* an AR model to an MA model that is infinite in extent and vice versa. It is often better practice to have a parsimonious model that combines the features of both an AR(p) and an MA(q) model. Such a model is then called an ARMA(p, q) model.

DEFINITION 1.106. An ARMA(p, q) model contains both an AR(p) and MA(q) portion

$$\begin{aligned} y_t = & \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} \\ & + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \end{aligned} \quad (1.141)$$

We can rewrite this model as

$$\begin{aligned} y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \cdots - \phi_p y_{t-p} = \\ \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \end{aligned}$$

or, using the backshift operator

$$\begin{aligned} (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) y_t = \delta + (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) \varepsilon_t \\ \phi(B) y_t = \delta + \theta(B) \varepsilon_t. \end{aligned} \quad (1.142)$$

THEOREM 1.107. *The ARMA(p, q) model is stationary if, and only if the roots of the characteristic equation*

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$$

are outside the unit circle and θ_i is square summable.

PROOF. If the roots of $\phi(z)$ are outside the unit circle we can invert (1.142) from Theorem 1.56 and write it as

$$y_t = \delta' + \psi'(B) \varepsilon_t$$

with

$$\begin{aligned} \psi'(z) &= \frac{1 - \theta_1 z - \theta_2 z^2 - \cdots - \theta_q z^q}{1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p} \\ &= 1 - \psi_1 z - \psi_2 z^2 - \psi_3 z^3 - \cdots \end{aligned}$$

and

$$\delta' = \frac{\delta}{1 - \phi_1 - \phi_2 - \cdots - \phi_p}.$$

Then

$$\sum_{i=0}^{\infty} |\psi_i| < \infty$$

from Theorem 1.93 only if the roots are outside the unit circle and if θ_i is square summable and the model is stationary. Note that if θ_i is finite or square summable the stationarity is entirely determined by the AR(p) component of the model. $\square$

THEOREM 1.108. *The mean of a stationary ARMA(p, q) model is*

$$\mu = \frac{\delta}{1 - \phi_1 - \phi_2 - \cdots - \phi_p}. \quad (1.143)$$

PROOF. Taking expectations on both sides of (1.141) gives the result. $\square$

THEOREM 1.109. *The autocovariance after lag q of a stationary ARMA(p, q) model follows*

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots + \phi_p \gamma_{k-p}, \quad k = q+1, q+2, \dots \quad (1.144)$$

PROOF. If we apply (1.143) to rewrite (1.141) as

$$\begin{aligned} y_t - \mu &= \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \cdots + \phi_p(y_{t-p} - \mu) \\ &\quad + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \end{aligned}$$

and then multiply both sides with $y_{t-k} - \mu$ and take expectations the result follows as in Theorem (1.87). $\square$

The first few autocovariances $k = 0, 1, \dots, q$ do not conform to (1.144). Here the autocovariance includes portions contributed from ε_t that falls away later on. As an example of calculating the first few autocovariances see Theorem 1.112.

COROLLARY 1.110. *The autocorrelation after lag q of a stationary ARMA(p, q) model also follows*

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}, \quad k = q+1, q+2, \dots \quad (1.145)$$

8.3.1. *Specification Problems.* The duality between AR and MA models necessitates care against *over parameterizing* an ARMA model (Hamilton 1994:60). If we for example have white noise

$$y_t = \varepsilon_t$$

but specify the model as a perfectly valid ARMA(1, 1) model

$$(1 - \phi_1 B)y_t = (1 - \theta_1 B)\varepsilon_t$$

the data will be described correctly for any $\phi_1 = \theta_1$. If we try to estimate ϕ_1 and θ_1 from a sample we will run into trouble and from a theoretical point of view miss out on a few important cancellations. We are better off modeling y_t as white noise rather than an ARMA(1, 1) model if we find that ϕ_1 is close to θ_1 .

In a similar fashion we can rewrite a stationary ARMA(p, q) model as

$$\begin{aligned} (1 - \lambda_1 B)(1 - \lambda_2 B) \dots (1 - \lambda_p B)(y_t - \mu) = \\ (1 - \kappa_1 B)(1 - \kappa_2 B) \dots (1 - \kappa_q B)\varepsilon_t \end{aligned} \quad (1.146)$$

with $|\lambda_i| < 1$. If any of these roots are the same, say

$$\lambda = \lambda_i = \kappa_j$$

then we can (and should!) divide with $(1 - \lambda B)$ to obtain an identical ARMA($p-1, q-1$) model.

8.3.2. *The ARMA(1, 1) Model.* The simplest (and in practice often most useful) ARMA model is the ARMA(1, 1) model

$$y_t = \delta + \phi_1 y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}. \quad (1.147)$$

THEOREM 1.111. *The mean of the stationary ARMA(1, 1) model is obtained from (1.143) as*

$$\mu = \frac{1}{1 - \phi_1}. \quad (1.148)$$

THEOREM 1.112. *The autocovariance of the stationary ARMA(1, 1) model is*

$$\gamma_k = \begin{cases} \sigma_\varepsilon^2 \cdot \frac{1-2\theta_1\phi_1+\theta_1^2}{1-\phi_1^2}, & k = 0 \\ \sigma_\varepsilon^2 \cdot \frac{(1-\theta_1\phi_1)(\phi_1-\theta_1)}{1-\phi_1^2}, & k = 1 \\ \phi_1\gamma_{k-1}, & k = 2, 3, \dots \end{cases} \quad (1.149)$$

PROOF. The result for $k \geq 2$ follows from (1.144). For $k = 0, 1$ we write the model as

$$(y_t - \mu) = \phi_1(y_{t-1} - \mu) + \varepsilon_t - \theta_1\varepsilon_{t-1}$$

and multiply this with $y_{t-k} - \mu$ and take expectations to obtain

$$\gamma_k = E[\phi_1(y_{t-1} - \mu)(y_{t-k} - \mu) + \varepsilon_t(y_{t-k} - \mu) - \theta_1\varepsilon_t(y_{t-k} - \mu)].$$

After two backward substitutions we have

$$\begin{aligned} \gamma_k &= \phi_1\gamma_{k-1} + E\left[\varepsilon_t(\phi_1^2 y_{t-k-2} + \varepsilon_{t-k} + (\phi_1 - \theta_1)\varepsilon_{t-k-1} \right. \\ &\quad \left. - \theta_1\varepsilon_{t-k-2} - \mu) - \theta_1\varepsilon_{t-1}(\phi_1^2 y_{t-k-2} + \varepsilon_{t-k} \right. \\ &\quad \left. + (\phi_1 - \theta_1)\varepsilon_{t-k-1} - \theta_1\varepsilon_{t-k-2} - \mu)\right] \\ &= \begin{cases} \phi_1\gamma_1 + \sigma_\varepsilon^2 - \theta_1(\phi_1 - \theta_1), & k = 0 \\ \phi_1\gamma_0 - \theta_1\sigma_\varepsilon^2, & k = 1 \end{cases}. \end{aligned}$$

Solving these equations for γ_0 and γ_1 yields

$$\gamma_1 = \phi_1^2\gamma_1 + \phi_1\sigma_\varepsilon^2 - \phi_1\theta_1(\phi_1 - \theta_1)\sigma_\varepsilon^2 - \theta_1\sigma_\varepsilon^2$$

$$\begin{aligned} \text{thus } \gamma_1 &= \sigma_\varepsilon^2 \cdot \frac{\phi_1 - \phi_1^2\theta_1 + \phi_1\theta_1^2 - \theta_1}{1 - \phi_1^2} \\ &= \sigma_\varepsilon^2 \cdot \frac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 - \phi_1^2} \end{aligned}$$

and similarly

$$\gamma_0 = \phi_1(\phi_1\gamma_0 - \theta_1\sigma_\varepsilon^2) + \sigma_\varepsilon^2 - \theta_1(\phi_1 - \theta_1)\sigma_\varepsilon^2$$

$$\text{or } \gamma_0 = \sigma_\varepsilon^2 \cdot \frac{1 + \theta_1^2 - 2\theta_1\phi_1}{1 - \phi_1^2}.$$

□

8.4. Model Identification. We saw that while the autocorrelation of an $MA(q)$ model cuts off after lag q the partial autocorrelation is infinite in extent. Similarly the autocorrelation of an $AR(p)$ model is infinite in extent but the partial autocorrelation cuts off after p lags. These properties assist us in selecting or identifying the appropriate $ARMA(p, q)$ model to use through selecting suitable values for p and q .

The approach is quite straightforward. Once the data has been transformed to a stationary time series we plot the *correlogram* or SAC and the *partial correlogram* or SPAC and examine these to determine one or more possible models. If the correlogram have only a few significant ‘spikes’ and then ‘cuts off’ we use the number of spikes as a starting point for the q parameter and similarly for the p parameter using the partial correlogram.

We can characterize the behaviour of the correlogram or autocorrelations of an $AR(p)$ model fairly accurately if we assume that the roots of the characteristic equation are distinct. From Corollary 1.97 the autocorrelations follow the same difference equation

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}$$

which can also be written as in (1.55) as

$$\boldsymbol{\rho}_k = \mathbf{F} \boldsymbol{\rho}_{k-1}. \quad (1.150)$$

This equation has the same characteristic equation and the same distinct roots as the model itself. If we apply the usual back substitution to (1.150) we obtain

$$\boldsymbol{\rho}_k = \mathbf{F}^n \boldsymbol{\rho}_{k-n}.$$

This process starts at $\rho_0 = 1$ and we can apply the back substitution until we get something similar to (1.57)

$$\rho_k = f_{11}^{(k-p)} \rho_p + f_{12}^{(k-p)} \rho_{p-1} + f_{13}^{(k-p)} \rho_{p-2} + \cdots + f_{1p}^{(k-p)} \rho_0. \quad (1.151)$$

Since the roots of the characteristic equation or equivalently, the eigenvalues of the matrix $\mathbf{F}$ are distinct we can apply (1.61) to write (1.151) as

$$\rho_k = \beta_1 \lambda_1^k + \beta_2 \lambda_2^k + \cdots + \beta_p \lambda_p^k \quad (1.152)$$

with λ_i the eigenvalues of the matrix $\mathbf{F}$ and $|\lambda_i| < 1$.

If an eigenvalue λ_i is real then $\beta_i \lambda_i^k$ in (1.152) decays exponentially to zero as k increases which gives rise to a damped exponential in the correlogram. On the other hand if λ_i is complex then there exists a λ_j that is its conjugate and from Corollary 1.53 it follows that β_i and β_j are also conjugates, say

$$\beta_i = a + \iota b,$$

$$\beta_j = a - \iota b.$$

Writing λ_i and λ_j in modulus argument form gives

$$\lambda_i^k = d^k (\cos(k\theta) + \iota \sin(k\theta))$$

$$\lambda_j^k = d^k (\cos(k\theta) - \iota \sin(k\theta))$$

so that, as in the proof of Theorem (1.84), we have

$$\begin{aligned} \beta_i \lambda_i^k + \beta_j \lambda_j^k &= 2d^k (a \cos(k\theta) - b \sin(k\theta)) \\ &= 2d^k P(k\theta) \\ &= 2d^k P(k\theta + 2\pi). \end{aligned}$$

Note that as indicated this P is a periodic function with period 2π . Since $d < 1$ the whole expression $2d^k P(k\theta)$ decays to zero as k increases. Thus if the characteristic equation has complex roots we will observe damped periodic waves in the correlogram.

As a simple example, consider the AR(2) model

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t.$$

with characteristic equation

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2.$$

The roots of this quadratic equation is given by

$$\lambda = -\frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2}. \quad (1.153)$$

These roots are either real and distinct, complex conjugates or real and equal.

<i>Model</i>	<i>Autocorrelation</i>	<i>Partial Autocorrelation</i>
AR(1)	decays exponentially (alternating signs if $\phi_1 < 0$)	single spike
MA(1)	single spike	decays exponentially (alternating signs if $\theta_1 < 0$)
AR(p)	decays exponentially, may contain damped oscillations	p spikes
MA(q)	q spikes	decays exponentially, may contain damped oscillations
ARMA(p, q)	both decay exponentially and may contain damped oscillations	

TABLE 1.9. Behaviour of ARMA Models used for Model Identification

1. The roots will be real and distinct if $\phi_1^2 + 4\phi_2 > 0$. The correlogram will then contain the sum of two damped exponentials.
2. The roots will be complex conjugates if $\phi_1^2 + 4\phi_2 < 0$. In this case the correlogram will contain damped sine and cosine waves.
3. The roots will be real and equal if $\phi_1^2 + 4\phi_2 = 0$. We at least know from the proof of Theorem 1.84 that the correlogram will decay to zero.

The partial autocorrelation of an MA(q) model is similar to the autocorrelation of an AR(p) model, and will also contain damped exponentials and damped sine and cosine waves (Box & Jenkins 1976:71).

8.5. Moments of ARMA Models. In this section we investigate higher order moments of a time series y_t that follows a general ARMA(p, q)

model. We have already derived the first and second order moments in calculating the mean and variance and now will specifically be interested in the fourth moment since this is used to calculate the *kurtosis* of the model. If this is known we can determine if the model is *leptokurtic* or *platykurtic* which has great implications on the *risk* of y_t if it represents the returns of an investment in some financial instrument.

DEFINITION 1.113. The r th moment about the mean of a random variable y is

$$\mu_r = E[y - \mu]^k. \quad (1.154)$$

and the r th moment about the origin is

$$\mu'_r = E[y^k] \quad (1.155)$$

COROLLARY 1.114. The first moment about the mean is zero.

COROLLARY 1.115. The first moment about the origin is the mean μ .

COROLLARY 1.116. The second moment about the mean is the variance γ_0 .

COROLLARY 1.117. The second moment about the origin is

$$\mu'_2 = \gamma_0 + \mu^2. \quad (1.156)$$

PROOF.

$$\begin{aligned} \gamma_0 &= E[y - \mu]^2 \\ &= E[y^2 - 2y\mu + \mu^2] \\ &= E[y^2] - 2\mu E[y] + \mu^2 \\ &= \mu'_2 - \mu^2. \end{aligned} \quad (1.157)$$

□

DEFINITION 1.118. The moment generating function of a stochastic variable y is the expectation

$$M_y(k) = E[e^{ky}] \quad (1.158)$$

provided that this expectation exists for all values of k in some open interval about the origin $-a < k < a$.

DEFINITION 1.119. The kurtosis of a distribution is defined as

$$\kappa = \frac{\mu_4}{\mu_2^2}.$$

If we have a stationary time series it is usually easier to transform it to a zero mean time series and to use this transformed time series to calculate higher order moments. As usual we will use z_t to denote a zero mean time series for which we will calculate *moments* rather than moments about the origin or moments about the mean, which is the same for z_t .

One good reason for introducing the moment generating function lies in its uniqueness. This gives rise to the well known result that if two variables have the same moment generating function then they also have the same cumulative probability distribution (Bain & Engelhardt 1986:202). This result allows us to identify the distribution that a variable follows and also is used to determine the distribution of the sum of independent variables from the property

$$e^{y_1+y_2} = e^{y_1} \cdot e^{y_2}. \quad (1.159)$$

Finally note that we can use the series expansion to write

$$\begin{aligned} M_y(k) &= E[e^{ky}] \\ &= E\left[1 + ky + \frac{k^2 y^2}{2} + \frac{k^3 y^3}{3!} + \dots\right] \end{aligned}$$

so that the expected value $E[y^r]$ is given by the coefficient of $\frac{k^r}{r!}$ in the moment generating function.

If z_t follows an ARMA(p, q) model we write

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (1.160)$$

and it follows that knowledge of the higher moments of z_t requires knowledge of the higher moments of ε_t . We have until now only assumed ε_t to be white noise but we will have to specify the distribution that it has as well in

order to be able to calculate these higher moments. The distribution that immediately comes to the fore is the normal distribution.

8.5.1. *Properties of the Normal Distribution.* The normal distribution was first published by Abraham de Moivre in 1733 as an approximation for the distribution of a sum of binomial random variables. Whether this is viewed as a blessing or curse, it has since become the single most important distribution in probability and statistics.

DEFINITION 1.120. A stochastic variable follows a normal distribution with mean μ and variance σ^2 , denoted as

$$y \sim N(\mu, \sigma^2) \quad (1.161)$$

if it has a *probability density function* given by

$$f(y, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} \quad (1.162)$$

and *cumulative probability density function*

$$F(y, \mu, \sigma) = \int_{-\infty}^y f(t, \mu, \sigma) dt. \quad (1.163)$$

DEFINITION 1.121. We call the normal distribution with $\mu = 0$ and $\sigma^2 = 1$ the *standard normal distribution* and if y follows this we write

$$y \sim N(0, 1). \quad (1.164)$$

This is of course well known to the reader. In order to investigate moments we will make use of some of the known properties of the normal distribution.

THEOREM 1.122. *The moment generating function of the standard normal distribution is*

$$M_z(k) = e^{\frac{k^2}{2}}. \quad (1.165)$$

PROOF. If $z \sim N(0, 1)$ we have

$$\begin{aligned} M_z(k) &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} e^{kz} dz \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-k)^2}{2} + \frac{k^2}{2}} dz \\ &= e^{\frac{k^2}{2}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-k)^2}{2}} dz \\ &= e^{\frac{k^2}{2}}. \end{aligned}$$

The final integral is over the density function of a normal variable and equals 1. □

THEOREM 1.123. *If y is distributed normally, $y \sim N(\mu, \sigma^2)$ then*

$$z = \frac{y - \mu}{\sigma} \sim N(0, 1) \quad (1.166)$$

$$M_y(k) = e^{\mu k + \frac{\sigma^2 k^2}{2}} \quad (1.167)$$

$$\mu_{2k} = E[y - \mu]^{2k} = \frac{(2k)! \sigma^{2k}}{k! 2^k} \quad (1.168)$$

$$\mu_{2k+1} = E[y - \mu]^{2k+1} = 0 \quad (1.169)$$

$$\kappa = \frac{\mu_4}{\mu_2^2} = 3. \quad (1.170)$$

PROOF. Most of the work is easy to do after making the transformation $z = \frac{y - \mu}{\sigma}$ and $\frac{dy}{dz} = \sigma$. It follows that

$$\begin{aligned} F_{(z)}(z, 0, 1) &= \int_{-\infty}^z f_{(z)}(t, 0, 1) dt \\ &= \int_{-\infty}^y \frac{f_{(z)}(\frac{u - \mu}{\sigma}, 0, 1)}{\sigma} du \\ &= \int_{-\infty}^y f_{(y)}(u, \mu, \sigma^2) du \\ &= F_{(y)}(y, \mu, \sigma^2) \end{aligned}$$

which yields (1.166). To show (1.167) we use Theorem 1.122 and write

$$\begin{aligned} M_y(k) &= M_{\sigma z + \mu}(k) \\ &= e^{\mu k} M_z(\sigma k) \\ &= e^{\mu k + \frac{\sigma^2 k^2}{2}}. \end{aligned}$$

Finally (1.168) and (1.169) follows from the series expansion

$$\begin{aligned} M_{y-\mu}(t) &= e^{\frac{\sigma^2 t^2}{2}} \\ &= \sum_{k=0}^{\infty} \frac{\sigma^{2k} t^{2k}}{2^k k!} \\ &= \sum_{k=0}^{\infty} \frac{\sigma^{2k} (2k)! t^{2k}}{2^k k! (2k)!} \end{aligned}$$

and then the r th moment is the coefficient of $\frac{k^r}{r!}$. Since only even powers of k appear the odd moments are zero. $\square$

The following lemma allows us to express (1.168) slightly differently.

THEOREM 1.124.

$$\prod_{i=1}^k (2i-1) = \frac{(2k)!}{k! 2^k} \quad k = 1, 2, 3, \dots \quad (1.171)$$

PROOF. First we simplify the two sides, say A and B , for k .

$$\begin{aligned} A_k &= \prod_{i=1}^k (2i-1) \\ &= 1 \cdot 3 \cdot 5 \cdot \dots \cdot (2k-1) \\ B_k &= \frac{(2k)!}{k! 2^k} \\ &= \frac{(2k) \cdot (2k-1) \cdot \dots \cdot k \cdot (k-1) \cdot \dots \cdot 1}{(2 \cdot 2 \cdot \dots \cdot 2)(k \cdot (k-1) \cdot \dots \cdot 1)} \\ &= \frac{k+1}{2} \cdot \frac{k+2}{2} \cdot \dots \cdot \frac{2k}{2} \end{aligned}$$

and, with $k = 1$, we have $A_1 = B_1$. If we assume $A_k = B_k$ then we have to show that $A_{k+1} = B_{k+1}$

$$\begin{aligned}
 A_{k+1} &= 1 \cdot 3 \cdot \dots \cdot (2k-1) \cdot (2(k+1)-1) \\
 &= A_k \cdot (2(k+1)-1) \\
 &= A_k \cdot (2k+1) \\
 B_{k+1} &= \frac{k+2}{2} \cdot \frac{k+3}{2} \cdot \dots \cdot \frac{2k}{2} \cdot \frac{2k+1}{2} \cdot \frac{2k+2}{2} \\
 &= (B_k \cdot \frac{2}{k+1}) \cdot \frac{2k+1}{2} \cdot \frac{2k+2}{2} \\
 &= B_k \cdot (2k+1).
 \end{aligned}$$

□

Finally we proof the following important theorem that states that a linear combination of normal variables again follows a normal distribution.

THEOREM 1.125. *If $y_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2, \dots, n$ then*

$$y' = \sum_{i=1}^n a_i y_i \sim N\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right). \quad (1.172)$$

PROOF.

$$\begin{aligned}
 M_{y'}(t) &= \prod_{i=1}^n \\
 &= \prod_{i=1}^n e^{a_i \mu_i t + a_i^2 t^2 \frac{\sigma_i^2}{2}} \\
 &= e^{t \sum_{i=1}^n a_i \mu_i + t^2 \sum_{i=1}^n a_i^2 \frac{\sigma_i^2}{2}}
 \end{aligned}$$

The final expression is the moment generating function of a normal distribution as in (1.172). □

8.6. The ARMA(1,1) Risk and Kurtosis. We know that the variance of an ARMA(1,1) model is provided by

$$\gamma_0 = \sigma_\varepsilon^2 \frac{1 + \theta_1^2 - 2\theta_1\phi_1}{1 - \phi_1^2}$$

from (1.149). If we assume σ_ε to be constant, say 1, and plot the resulting contours of γ_0 we obtain Figure 1.11. Note that the variance of this model

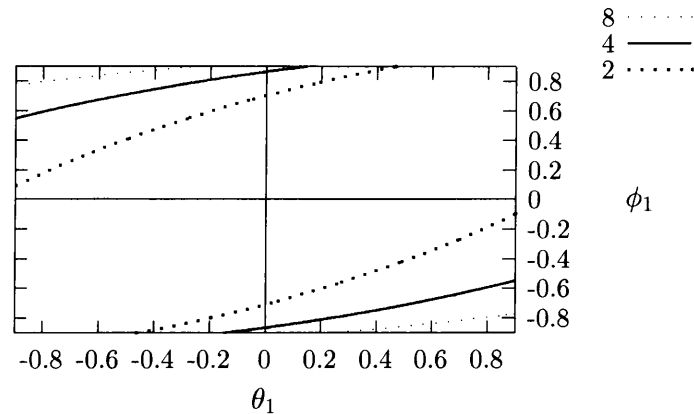


FIGURE 1.11. Contours of ARMA(1,1) Variance

is much more sensitive to changes in ϕ_1 than changes in θ_1 . If ϕ_1 is thus high in magnitude we know that the underlying e.g. asset is risky and also that our predictions using this model will not be that accurate.

We also need to look at the kurtosis to comment on the accuracy of predictions. If we rewrite the model through backsubstitution

$$\begin{aligned}
 z_t &= \phi_1 z_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} \\
 &= \phi_1 (\phi_1 z_{t-2} + \varepsilon_{t-1} - \theta_1 \varepsilon_{t-2}) + \varepsilon_t - \theta_1 \varepsilon_{t-1} \\
 &= \phi_1^2 z_{t-2} + \varepsilon_t + (\phi_1 - \theta_1) \varepsilon_{t-1} - \phi_1 \theta_1 \varepsilon_{t-2} \\
 &\vdots \\
 &= \phi_1^k z_{t-k} + \varepsilon_t + \sum_{i=1}^k (\phi_1^{k-i} (\phi_1 - \theta_1) \varepsilon_{t-i}) \\
 &= \left(1 + \sum_{i=1}^{\infty} \phi_1^{k-i} (\phi_1 - \theta_1) B^i \right) \varepsilon_t
 \end{aligned}$$

if the model is invertible. This expression is an infinite linear combination of ε_t and, if we assume the errors to be normally distributed, results in another

normal distribution from (1.172) so that

$$z_t \sim N(0, \gamma_0). \quad (1.173)$$

Note that this will hold for any stationary and invertible ARMA(p, q) model that we can rewrite as an infinite sum of the errors. Thus we can calculate all the higher moments of the model if we assume that the errors are normally distributed.

THEOREM 1.126. *The stationary and invertible ARMA(p, q) model with normal errors itself follows a normal distribution*

$$z_t \sim N(0, \gamma_0).$$

As long as we are able to determine the distribution of a linear combination of the errors of the ARMA(p, q) model we will be able to determine the distribution that the actual model follows. If we are not able to determine this distribution or if it is not known we can still obtain an expression for e.g. the kurtosis through direct calculation. We are often interested in the kurtosis in order to comment further on the inherent risk in the model.

In order to determine the kurtosis we need to calculate the fourth moment of z_t . If z_t follows an ARMA(1,1) model this moment is

$$E[z_t^4] = E[(z_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1})^4].$$

If we develop the right hand side we obtain 15 terms that we need to solve. In solving these terms we will often make use of the fact that ε_t is i.i.d. with mean zero. Note that if two variables are independent then so are their squares and other higher powers. If two variables, say z_{t-1} and ε_t are independent then it is well known that

$$E[z_{t-1} \varepsilon_t] = E[z_{t-1}] E[\varepsilon_t].$$

The following expressions will also be used to evaluate these 15 terms:

$$E[\varepsilon_t^3] = \mu_3 \quad (1.174)$$

$$E[\varepsilon_t^4] = \mu_4 \quad (1.175)$$

$$E[z_t^2] = \gamma_0 = \sigma_\varepsilon^2 \frac{1 + \theta_1^2 - 2\theta_1\phi_1}{1 - \phi_1^2} \quad \text{from (1.149)} \quad (1.176)$$

$$E[z_t \varepsilon_t] = E[\phi_1 z_{t-1} \varepsilon_t + \varepsilon_t^2 - \theta_1 \varepsilon_{t-1} \varepsilon_t] = \sigma_\varepsilon^2 \quad (1.177)$$

$$E[z_t \varepsilon_t^2] = E[\phi_1 z_{t-1} \varepsilon_t^2 + \varepsilon_t^3 - \theta_1 \varepsilon_{t-1} \varepsilon_t^2] = 0 \quad (1.178)$$

$$E[z_t \varepsilon_t^3] = E[\phi_1 z_{t-1} \varepsilon_t^3 + \varepsilon_t^4 - \theta_1 \varepsilon_{t-1} \varepsilon_t^3] = \mu_4 \quad (1.179)$$

$$\begin{aligned} E[z_{t-i} \varepsilon_t^k] &= E[\phi_1 z_{t-i-1} \varepsilon_t^k + \varepsilon_{t-i} \varepsilon_t^k - \theta_1 \varepsilon_{t-i-1} \varepsilon_t^k] \\ &= 0 \quad i = 1, 2, \dots \end{aligned} \quad (1.180)$$

We also need expressions for $E[z_t^3 \varepsilon_t]$ and $E[z_t^2 \varepsilon_t^2]$. If we expand the powers and take expectations we obtain

$$\begin{aligned} E[z_t^3 \varepsilon_t] &= E[(\phi_1^3 z_{t-1}^3 + 3\phi_1^2 z_{t-1}^2 \varepsilon_t - 3\phi_1^2 \theta_1 z_{t-1}^2 \varepsilon_{t-1} + 3\phi_1 z_{t-1} \varepsilon_t^2 \\ &\quad - 6\phi_1 \theta_1 z_{t-1} \varepsilon_{t-1} \varepsilon_t + 3\phi_1 \theta_1^2 z_{t-1} \varepsilon_{t-1}^2 + \varepsilon_t^3 - 3\theta_1 \varepsilon_{t-1} \varepsilon_t^2 \\ &\quad + 3\theta_1^2 \varepsilon_{t-1}^2 \varepsilon_t - \theta_1^3 \varepsilon_{t-1}^3) \cdot \varepsilon_t] \\ &= 3\phi_1^2 \gamma_0 \sigma_\varepsilon^2 - 6\phi_1 \theta_1 \sigma_\varepsilon^4 + \mu_4 + 3\theta_1^2 \sigma_\varepsilon^4 \end{aligned} \quad (1.181)$$

$$\begin{aligned} E[z_t^2 \varepsilon_t^2] &= E[(\phi_1^2 z_{t-1}^2 + 2\phi_1 z_{t-1} \varepsilon_t - 2\phi_1 \theta_1 z_{t-1} \varepsilon_{t-1} + \varepsilon_t^2 \\ &\quad - 2\theta_1 \varepsilon_{t-1} \varepsilon_t + \theta_1^2 \varepsilon_{t-1}^2) \cdot \varepsilon_t^2] \\ &= \phi_1^2 \gamma_0 \sigma_\varepsilon^2 - 2\phi_1 \theta_1 \sigma_\varepsilon^4 + \mu_4 + \theta_1^2 \sigma_\varepsilon^4. \end{aligned} \quad (1.182)$$

Now we have the tools we need to evaluate the fourth moment. The 15 terms as well as their resulting expressions are given in Table 1.10.

The next step is to add these 15 terms and simplify the resulting expression. After some work we obtain the result as

$$E[z_t^4] = 6 \cdot \frac{(\phi_1 - \theta_1)^2 (\phi_1^2 \theta_1^2 - 2\phi_1^3 \theta_1 + 1)}{(1 - \phi_1^2)(1 - \phi_1^4)} \cdot \sigma_\varepsilon^4 \quad (1.183)$$

$$+ \frac{\theta_1^4 - 4\phi_1 \theta_1^3 + 6\phi_1^2 \theta_1^2 - 4\phi_1^3 \theta_1 + 1}{1 - \phi_1^4} \cdot \mu_4. \quad (1.184)$$

Term	Expectation	Reason
$\phi_1^4 z_{t-1}^4 = \phi_1^4 E[z_t^4]$	same unconditional expectation	
$\theta_1^4 \varepsilon_{t-1}^4 = \theta_1^4 \mu_4$		by (1.175)
$\varepsilon_t^4 = \mu_4$		by (1.175)
$4\phi_1 z_{t-1} \varepsilon_t^3 = 0$		$E[z_{t-1}] = 0$
$-4\phi_1 \theta_1^3 z_{t-1} \varepsilon_{t-1}^3 = -4\phi_1 \theta_1^3 \mu_4$		by (1.179)
$12\phi_1 \theta_1^2 z_{t-1} \varepsilon_{t-1}^2 \varepsilon_t = 0$		$E[\varepsilon_t] = 0$
$-12\phi_1 \theta_1 z_{t-1} \varepsilon_{t-1} \varepsilon_t^2 = -12\phi_1 \theta_1 \sigma_\varepsilon^4$		by (1.177)
$-12\phi_1^2 \theta_1 z_{t-1}^2 \varepsilon_{t-1} \varepsilon_t = 0$		$E[\varepsilon_t] = 0$
$6\phi_1^2 z_{t-1}^2 \varepsilon_t^2 = 6\phi_1^2 \gamma_0 \sigma_\varepsilon^2$		by (1.176)
$6\phi_1^2 \theta_1^2 z_{t-1}^2 \varepsilon_{t-1}^2 = 6\phi_1^2 \theta_1^2 E[z_t^2 \varepsilon_t^2]$		we will use (1.182)
$4\phi_1^3 z_{t-1}^3 \varepsilon_t = 0$		$E[\varepsilon_t] = 0$
$-4\phi_1^3 \theta_1 z_{t-1}^3 \varepsilon_{t-1} = -4\phi_1^3 \theta_1 E[z_t^3 \varepsilon_t]$		we will use (1.181)
$-4\theta_1 \varepsilon_{t-1} \varepsilon_t^3 = 0$		$E[\varepsilon_{t-1}] = 0$
$-4\theta_1^3 \varepsilon_{t-1}^3 \varepsilon_t = 0$		$E[\varepsilon_t] = 0$
$6\theta_1^2 \varepsilon_{t-1}^2 \varepsilon_t^2 = 6\theta_1^2 \sigma_\varepsilon^4$		$E[\varepsilon_t^2] = \sigma_\varepsilon^2$

TABLE 1.10. Expected Terms of Fourth Moment of
ARMA(1,1) Model

If we define the kurtosis of the error ε_t as

$$\kappa_\varepsilon = \frac{\mu_4}{\sigma_\varepsilon^2} \quad (1.185)$$

we can calculate the kurtosis of the ARMA(1,1) model or z_t as

$$\begin{aligned}
 \kappa &= \frac{E[z_t^4]}{\gamma_0^2} \\
 &= \frac{4\phi_1^5\theta_1 - 6\phi_1^4\theta_1^2 + 4\phi_1^3\theta_1(\theta_1^2 - 1) - \phi_1^2(\theta_1^4 - 6\theta_1^2 + 1) - 4\phi_1\theta_1^3 + \theta_1^4 + 1}{(1 + \phi_1^2)(\theta_1^2 - 2\phi_1\theta_1 + 1)^2} \cdot \kappa_\varepsilon \\
 &\quad - \frac{1}{6} \cdot \frac{2\phi_1^5\theta_1 - 5\phi_1^4\theta_1^2 + 4\phi_1^3\theta_1^3 - \phi_1^2(\theta_1^4 + 1) + 2\phi_1\theta_1 - \theta_1^2}{(1 + \phi_1^2)(\theta_1^2 - 2\phi_1\theta_1 + 1)^2} \\
 &= \alpha_1 \cdot \kappa_\varepsilon + \alpha_0.
 \end{aligned}$$

Thus the kurtosis of the ARMA(1,1) model is a linear function of the kurtosis of the errors. We already know that if the errors are normal so that $\kappa_\varepsilon = 3$ then we will have $\kappa = 3$. It is easy to see that this will hold by expanding the expression and grouping the terms in the numerator and denominator together as in Table 1.11. Notice in Table 1.11 how each term in the numerator is three times the corresponding term in the denominator if $\kappa_\varepsilon = 3$.

The kurtosis of the ARMA(1,1) model is a linear function of the kurtosis of the errors that always passes through the point (3,3). Since we have this point on the line available we only need either the slope or the intercept to complete the line's equation. Both the slope and the intercept are functions of ϕ_1 and θ_1 . The contours of the slope for various ϕ_1 and θ_1 are displayed in Figure 1.12.

Notice from Figure 1.12 that κ is more sensitive to changes in κ_ε when ϕ_1 and θ_1 have the same sign. The bigger the difference between ϕ_1 and θ_1 the less the slope α_1 and the less sensitive κ is to changes in κ_ε . Since the slope is less than 1 it also follows that the ARMA(1,1) model has a moderating effect on the kurtosis in the errors. If $\kappa_\varepsilon < 3$ so that ε_t is platykurtic then, since $0 < \alpha_1 < 1$ and the line passes through (3,3), it follows that $\kappa_\varepsilon < \kappa < 3$. Similarly we have $3 < \kappa < \kappa_\varepsilon$ if ε_t is leptokurtic or $\kappa_\varepsilon > 3$.

9. ARIMA Models

DEFINITION 1.127. The class of non stationary models known as *integrated models* is the class of models that may be made stationary through

Term in Numerator	Corresponding Term in Denominator
$+6(5 - \kappa_\varepsilon)\phi_1^4\theta_1^2$	$+4\phi_1^4\theta_1^2$
$-4(6 - \kappa_\varepsilon)\phi_1^3\theta_1^3$	$-4\phi_1^3\theta_1^3$
$-4\kappa_\varepsilon\phi_1^3\theta_1$	$-4\phi_1^3\theta_1$
$+(6 - \kappa_\varepsilon)\phi_1^2\theta_1^4$	$+\phi_1^2\theta_1^4$
$+6\kappa_\varepsilon\phi_1^2\theta_1^2$	$+6\phi_1^2\theta_1^2$
$+(6 - \kappa_\varepsilon)\phi_1^2$	$+\phi_1^2$
$-4\kappa_\varepsilon\phi_1\theta_1^3$	$-4\phi_1\theta_1^3$
$+\kappa_\varepsilon\theta_1^4$	$+\theta_1^4$
$+\kappa_\varepsilon$	$+1$
$-12\phi_1\theta_1$	$-4\phi_1\theta_1$
$+6\theta_1^2$	$+2\theta_1^2$
$-4(3 - \kappa_\varepsilon)\phi_1^5\theta_1$	

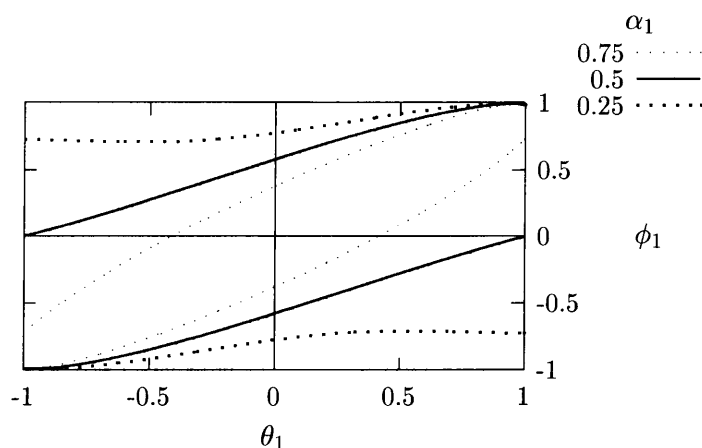
TABLE 1.11. Terms in Numerator and Denominator of
ARMA(1,1) Kurtosis

the application of a time-invariant filter or transformation function (Granger & Newbold 1986:41) such as

$$\begin{aligned}\tilde{y}_t &= y_t - y_{t-1} \\ &= (1 - B)y_t.\end{aligned}$$

We have seen that an ARMA model will be stationary if the roots of the characteristic equation are outside the unit circle. If the roots are inside the unit circle the model will explode and be non stationary. Consider the case where we have a simple model

$$(1 - B)y_t = \varepsilon_t$$

FIGURE 1.12. Contours of Slope α_1 in $\kappa = \alpha_0 + \alpha_1\kappa_\varepsilon$

so that

$$y_t = y_{t-1} + \varepsilon_t.$$

Here the model has just one root that lies on the unit circle. Note that this model is the random walk model (with no drift) considered earlier.

After taking first differences of this non stationary time series we obtain (stationary) white noise. We can in general obtain a stationary model after taking d differences of a non stationary model with d roots on and the rest outside the unit circle. In this case we can express the model as

$$\phi(B)y_t = \theta(B)\varepsilon_t$$

$$\varphi(B)(1-B)^d y_t = \theta(B)\varepsilon_t.$$

so that $(1-B)^d y_t$ follows a stationary ARMA(p, q) model.

DEFINITION 1.128. If $(1-B)^d y_t$ can be modeled as a stationary ARMA(p, q) model we say that y_t follows an ARIMA(p, d, q) or *autoregressive integrated moving average* model.

Although untransformed financial time series are normally not stationary financial returns are generally considered to be stationary. Note that if

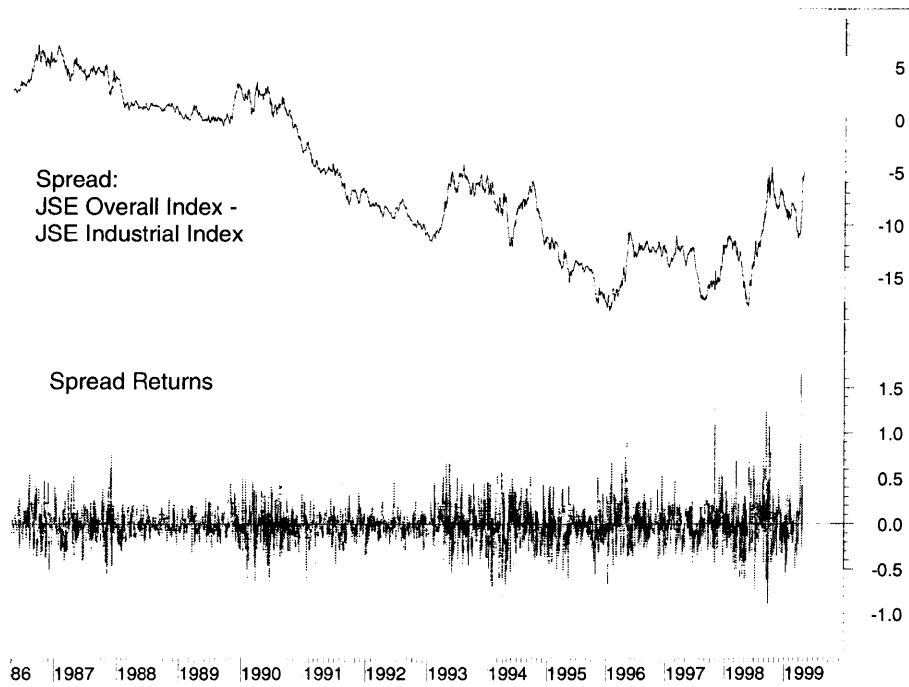


FIGURE 1.13. Example Spread and Spread Returns

we fit an $\text{ARMA}(p, q)$ model to specifically the log returns

$$\tilde{y}_t = \ln y_t - \ln y_{t-1}$$

we are implicitly assuming that $\ln y_t$ follows an $\text{ARIMA}(p, 1, q)$ model.

9.0.1. *Spreads.* Another frequent implicit application of ARIMA models is found when trading *spreads*. If a trader simultaneously buys y and sells x and later simultaneously sells off the y and buys back the x it is called a spread. This normally applies to futures on stock indices. The daily returns of the spread is

$$y_t - y_{t-1} - (x_t - x_{t-1}) = y_t - x_t - (y_{t-1} - x_{t-1})$$

to which an $\text{ARMA}(p, q)$ model may be fitted as an analysis tool. This implies the fitting of an $\text{ARIMA}(p, 1, q)$ model to $y_t - x_t$.

CHAPTER 2

Heteroscedastic Time Series Models

1. Introduction

All stationary time series models are *homoscedastic* by definition and most financial time series models in practice operated under this assumption.

DEFINITION 2.1. A time series is homoscedastic if it has constant variance.

DEFINITION 2.2. A heteroscedastic time series has non constant variance. Similarly a heteroscedastic time series model does not assume constant variance, but models it as a function of e.g. time.

Non stationarity usually refers to the relaxation of the constant mean assumption but it also refers to *heteroscedasticity*.

In the previous chapter we noted one example that would require a heteroscedastic time series model. In Figure 1.3 we saw how, after applying a transformation to the JSE All Share index, we obtain a time series with apparent constant mean but not constant variance. Heteroscedastic time series models cater for this non constant variance.

In the past the problem of heteroscedasticity was generally associated with cross-sectional rather than time series data (Gujarati 1988:319). Recently increasing emphasis has been placed on heteroscedastic financial time series (Engle 1993:72) and a number of new ways to model this variance or *volatility* have been developed. This new approach is well captured in Engle's (1993:72) remark: "*Financial market volatility is predictable.*"

1.1. Transformations for Heteroscedastic Time Series. As with the problem of a varying mean over time, the traditional approach to heteroscedasticity has been to transform it away using a suitable transformation. The most popular transformation applied is to simply take the logarithmic of the time series (Granger & Newbold 1986:308) – a transformation that is closely related to the calculation of returns. There have also been a lot of interest in a class of *variance stabilizing transformations* that was introduced by Box and Cox (1964):

$$y_t^{(\lambda)} = \begin{cases} \frac{y_t^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln y_t & \lambda = 0 \end{cases}.$$

Note that we have by l'Hôpital's rule

$$\lim_{\lambda \rightarrow 0} \frac{y_t^\lambda - 1}{\lambda} = \ln y_t.$$

This useful family of transformations is called the *power transformations* or the *Box-Cox transformations* after its inventors.

We typically have $-1 \leq \lambda \leq 1$ and (necessarily) $y_t > 0$. With $\mu_t = E[y_t]$ and provided that $\mu_t^2 \gg \lambda_0$ an approximation for the variance of the transformed variable is (Kendall & Ord 1990:54):

$$\text{Var} \left(y_t^{(\lambda)} \right) \approx \mu_t^{2\lambda-2} \text{Var}(y_t).$$

If the variance of y_t increases linearly with the mean we thus set $\lambda = \frac{1}{2}$ and if quadratically we set $\lambda = 0$.

It is often convenient to compare the standard deviation σ_t rather than the variance σ_t^2 to the mean as in Table 2.1. When estimating λ it normally suffices to perform a graphical check of the variance in the time series for the values $\lambda = -1, -\frac{1}{2}, 0, \frac{1}{2}, 1$ (Kendall & Ord 1990:54). Box and Cox (1964) provided a way to estimate λ using the method of maximum likelihood and the assumption that $\varepsilon_t \sim N(0, 1)$ for the proper value of λ .

If it is known or appears that the standard deviation of the untransformed time series y_t is a particular function of the mean, say $f(\mu_t)$, we can obtain an appropriate transformation $T(y_t)$ through

$$\frac{\partial T(y_t)}{\partial y_t} \propto \frac{1}{f(y_t)}.$$

	Nature of Dependence $\sigma_t = f(\mu_t)$		Variance Stabilizing Transformation
	$\sigma_t \propto \mu_t^k$	$y_t > 0$	y_t^{1-k}
	<i>and in particular</i>		
Poisson	$\sigma_t \propto \sqrt{\mu_t}$	$y_t > 0$	$\sqrt{y_t}$
	$\sigma_t \propto \mu_t$	$y_t > 0$	$\ln y_t$
	$\sigma_t \propto \mu_t^2$	$y_t > 0$	$\frac{1}{y_t}$
Binomial	$\sigma_t \propto \sqrt{\mu_t(1-\mu_t)}$	$0 < y_t < 1$	$\text{asin}(\sqrt{y_t})$
Negative binomial	$\sigma_t \propto \frac{\sqrt{1-\mu_t}}{\mu_t}$	$0 < y_t < 1$	$\sqrt{1-y_t} - \frac{\sqrt{(1-y_t)^3}}{3}$
	$\sigma_t \propto \frac{1}{(1-\mu_t^2)^2}$	$-1 < y_t < 1$	$\ln \frac{1+y_t}{1-y_t}$

TABLE 2.1. Appropriate Variance Stabilizing Transformations When $\sigma \propto f(\mu)$

Here we obtain $T(y_t)$ by integrating $\frac{1}{f(y_t)}$ with respect to y_t . Some well known transformations which arise in this way are shown in Table 2.1 (Drap-
er & Smith 1980:238).

Financial time series such as asset prices often increase exponentially over time. In this case $\sigma_t \propto \mu_t$ and the transformation $T = \ln y_t$ will be appropriate. Note for example the very clear linear relationship $\sigma_t \propto \mu_t$ between daily closes in the ZAR/USD exchange rate and a long term (1,000 days or roughly 4 years) rolling standard deviation in Figure 2.1. After taking logs the resulting time series still exhibits a long term trend that has been transformed from an exponential to a linear trend. The next step is to remove this trend and if we apply differencing

$$T(y_t) = \ln y_t - \ln y_{t-1}$$

we should end up with a fairly stationary time series. This transformation is equivalent to the log returns transformation discussed in Section 2.3.

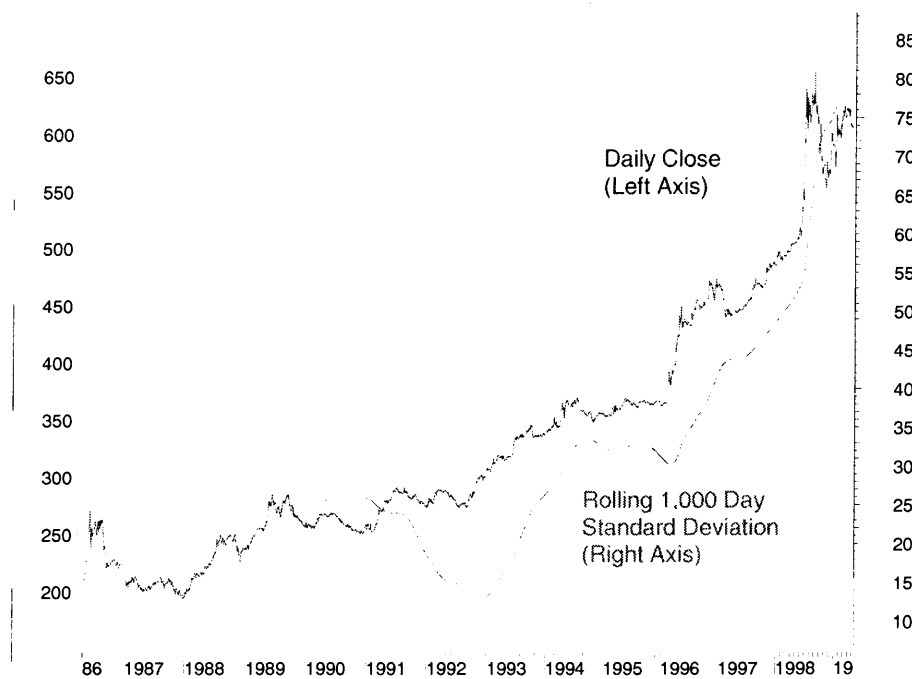


FIGURE 2.1. ZAR/USR Exchange Rate

2. ARCH Models

One of the most important heteroscedastic time series models was introduced less than twenty years ago by Engle (1982) as *autoregressive conditional heteroscedastic* or ARCH models. Later on Bollerslev (1986) generalized ARCH models into the *generalized ARCH* or GARCH models. These models rapidly gained in popularity and are seen by many as state of the art financial time series modeling. A recent survey conducted by Bollerslev (1992) found ARCH-related models cited in more than 200 papers on financial time series. Most modern risk management and econometric packages now cater in some way for these models.

When working with these models we will often make use of the *information set*:

DEFINITION 2.3. The set of available information at time t is defined as

$$\Psi_t. \quad (2.1)$$

Note that Ψ_t will contain as a subset all the y_k and ε_k with $k \leq t$ as well as all the model parameters.

2.1. ARCH Models.

DEFINITION 2.4. The heteroscedastic ARCH(p) model of order p is defined as:

$$\begin{aligned} y_t &= \varepsilon_t \sqrt{h_t} \\ h_t &= \alpha_0 + \alpha_1 y_{t-1}^2 + \alpha_2 y_{t-2}^2 + \cdots + \alpha_p y_{t-p}^2 \end{aligned} \quad (2.2)$$

with ε_t white noise and, importantly, constrained so that

$$\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2 = 1. \quad (2.3)$$

The error is usually assumed to be normal and then, using the information set, we can express (2.2) conveniently as

$$\begin{aligned} y_t | \Psi_{t-1} &\sim N(0, h_t) \\ h_t &= h(y_{t-1}, y_{t-2}, \dots, y_{t-p}, \boldsymbol{\alpha}). \\ &= h(\Psi_{t-1}, \boldsymbol{\alpha}) \end{aligned} \quad (2.4)$$

Note that the mean of this model is zero while the variance changes over time. Once we have variance estimates available we can also utilize it in other models. In financial terms we are busy forecasting volatility – an important concept in derivative pricing, risk management and asset allocation, amongst others. Engle (1982) estimated the volatility of inflation and later on this was related to other key macro economic variables (Coulson & Robins 1985:71-75). Weiss (1984) successfully modeled thirteen macro economic time series in the United States using ARMA models with ARCH errors.

2.2. The ARCH(1) Model.

DEFINITION 2.5. The first order ARCH(1) model is defined as

$$\begin{aligned} y_t &= \varepsilon_t \sqrt{h_t} \\ h_t &= \alpha_0 + \alpha_1 y_{t-1}^2 \end{aligned} \quad (2.5)$$

with $\alpha_0 > 0$ and $\alpha_1 \geq 0$ and $\sigma_\varepsilon^2 = 1$.

The restrictions we place on α_0 and α_1 are to ensure basic sanity when taking the root of h_t .

THEOREM 2.6. *The odd moments of an ARCH(1) model are zero.*

PROOF. The mean or first moment of the ARCH(1) model is zero

$$E[y_t] = E[\varepsilon_t \sqrt{h_t}] = E[\varepsilon_t] \sqrt{h_t} = 0.$$

That all other odd moments are zero follows from the symmetry of the distribution

$$E[y_t^{2k+1}] = \int_{-\infty}^{+\infty} y_t^{2k+1} f(y_t) dy_t. \quad (2.6)$$

Since y_t^{2k+1} is an odd function, $h(y) = -h(-y)$, multiplied by a symmetric function $f(y) = f(-y)$ it results in another odd function which always yields zero if integrated from $-b$ to $+b$. $\square$

THEOREM 2.7. *The conditional variance of an ARCH(1) model is*

$$\text{Var}(y_t | \Psi_{t-1}) = \alpha_0 + \alpha_1 y_{t-1}^2. \quad (2.7)$$

PROOF.

$$\begin{aligned} \text{Var}(y_t | \Psi_{t-1}) &= \text{Var}(\varepsilon_t \sqrt{h_t} | \Psi_{t-1}) \\ &= h_t \text{Var}(\varepsilon_t | \Psi_{t-1}) \\ &= \alpha_0 + \alpha_1 y_{t-1}^2. \end{aligned}$$

$\square$

A large observation y_{t-1} will lead to a large variance $\text{Var}(y_t)$ but the memory is confined to one period. *Volatility clustering* as noted in Figure 1.3 is one of the oldest noted characteristics of financial data (Engle 1993:72). If large changes in asset prices tend to be followed by more large changes it suggests a method of predicting future volatility following such changes. Traders typically make use of a suitable weighted moving average of previously observed volatilities as a proxy – as such utilizing the clustering phenomena.

If we assume that the errors are normal we can calculate the higher order even moments.

THEOREM 2.8. *The $2r$ th moment, $r = 0, 1, \dots$ of an ARCH(1) model with conditional normal distribution, $\alpha_0 > 0$ and $\alpha_1 \geq 0$, exists if, and only if*

$$\alpha_1^2 \prod_{i=1}^r (2i - 1) < 1.$$

PROOF. We define $\mathbf{w}_t$ as

$$\mathbf{w}_t = \begin{pmatrix} y_t^{2r} \\ y_t^{2(r-1)} \\ \vdots \\ y_t^2 \end{pmatrix} \quad (2.8)$$

and then by (1.168)

$$\begin{aligned} E[y_t^{2k} | \Psi_{t-1}] &= \frac{(2k)! \sigma^{2k}}{k! 2^k} \\ &= (\alpha_0 + \alpha_1 y_{t-1}^2)^k \frac{(2k)!}{k! 2^k} \\ &= \left[\sum_{i=0}^k \binom{k}{i} (\alpha_1 y_{t-1}^2)^{k-i} \alpha_0^i \right] c \\ &= \text{some linear combination of } \mathbf{w}_{t-1} \end{aligned}$$

$$\text{i.e.} \quad E[\mathbf{w}_t | \Psi_{t-1}] = \mathbf{b} + \mathbf{A} \mathbf{w}_{t-1} \quad (2.9)$$

where we can write $\mathbf{A}$ as a triangular matrix. Through backsubstituting we obtain

$$E[\mathbf{w}_t | \Psi_{t-2}] = \mathbf{b} + \mathbf{A}(\mathbf{b} + \mathbf{A} \mathbf{w}_{t-2})$$

and in general

$$E[\mathbf{w}_t | \Psi_{t-k}] = (\mathcal{I} + \mathbf{A} + \mathbf{A}^2 + \dots + \mathbf{A}^{k-1}) \mathbf{b} + \mathbf{A}^k \mathbf{w}_{t-k}. \quad (2.10)$$

If k increases without bound we have

$$\lim_{k \rightarrow \infty} E[\mathbf{w}_t | \Psi_{t-k}] = (\mathcal{I} - \mathbf{A})^{-1} \mathbf{b} \quad (2.11)$$

if, and only if all the eigenvalues of $\mathbf{A}$ are within the unit circle. As (2.11) does not depend on either t or $\mathbf{w}_{t-k}$ this limit gives the unconditional even moments of y_t

$$E[\mathbf{w}_t] = (\mathcal{I} - \mathbf{A})^{-1} \mathbf{b}.$$

The eigenvalues of $\mathbf{A}$ are the diagonal elements or the coefficients of y_t^{2k} in (2.9)

$$\begin{aligned} \lambda_k &= a_{k,k} \\ &= \alpha_1^k \frac{2k!}{k!2^k} \\ &= \alpha_1^k \prod_{i=1}^k (2i-1) \quad \text{by (1.171)} \\ &= \prod_{i=1}^k \alpha_1 (2i-1) \\ &= (\alpha_1 \cdot 1) \cdot (\alpha_1 \cdot 3) \cdot (\alpha_1 \cdot 5) \cdot \dots \cdot (\alpha_1 \cdot (2k-1)). \end{aligned}$$

Now λ_k is the product of k monotonically increasing and positive factors

$$\alpha_1 \cdot 1 < \alpha_1 \cdot 3 < \dots < \alpha_1 \cdot (2k-1).$$

If $\alpha_1 \cdot (2k-1) \leq 1$ then all the other factors will be less than 1 and $\lambda_{k-1} < 1$.

On the other hand, with $\alpha_1 \cdot (2k-1) > 1$ and

$$\lambda_{k-1} \cdot (\alpha_1 \cdot (2k-1)) = \lambda_k$$

it follows that $\lambda_{k-1} < \lambda_k$. Thus if $\lambda_r < 1$ we have all other eigenvalues $\lambda_k < 1$ for $k < r$. $\square$

For a finite second moment or variance we require that $\alpha_1 < 1$ and for a finite fourth moment we also require that $3\alpha_1^2 < 1$. Since these moments do not depend on time but only on the parameters of the model we can immediately give the stationary conditions of the model.

COROLLARY 2.9. *The ARCH(1) model is stationary if, and only if,*

$$\alpha_1 < 1. \quad (2.12)$$

Theorem 2.8 provides a method for calculating both conditional and unconditional ARCH(1) moments. Using (1.168) we have for example

$$E[y_t^{2k}|\Psi_{t-1}] = (\alpha_0 + \alpha_1 y_{t-1}^2)^k \prod_{i=1}^k (2i - 1)$$

$$E[y_t^2|\Psi_{t-1}] = \alpha_0 + \alpha_1 y_{t-1}^2$$

$$E[y_t^4|\Psi_{t-1}] = 3\alpha_0^2 + 6\alpha_0\alpha_1 y_{t-1}^2 + \alpha_1^2 y_{t-1}^4.$$

and, if we define

$$\mathbf{w}_t = \begin{pmatrix} y_t^4 \\ y_t^2 \end{pmatrix}$$

we can construct $\mathbf{A}$ and $\mathbf{b}$ as

$$\begin{aligned} E[\mathbf{w}_t|\psi_{t-1}] &= \begin{pmatrix} 3\alpha_0^2 \\ \alpha_0 \end{pmatrix} + \begin{pmatrix} 3\alpha_1^2 & 6\alpha_0\alpha_1 \\ 0 & \alpha_1 \end{pmatrix} \begin{pmatrix} y_{t-1}^4 \\ y_{t-1}^2 \end{pmatrix} \\ &= \mathbf{b} + \mathbf{A}\mathbf{w}_{t-1} \end{aligned}$$

and, after some work, obtain the unconditional

$$\begin{aligned} E[\mathbf{w}_t] &= (\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} \\ &= \begin{pmatrix} \frac{3\alpha_0^2}{(1-\alpha_1)^2} \cdot \frac{1-\alpha_1^2}{1-3\alpha_1^2} \\ \frac{\alpha_0}{1-\alpha_1} \end{pmatrix}. \end{aligned} \quad (2.13)$$

From (2.13) we obtain the unconditional fourth moment and variance of an ARCH(1) model as the first and second elements respectively.

COROLLARY 2.10. *The kurtosis of an ARCH(1) model is*

$$\kappa = \frac{3\alpha_0^2}{(1-\alpha_1)^2} \cdot \frac{1-\alpha_1^2}{1-3\alpha_1^2} \cdot \frac{(1-\alpha_1)^2}{\alpha_0^2} = 3 \cdot \frac{1-\alpha_1^2}{1-3\alpha_1^2}. \quad (2.14)$$

If $\alpha_1 = 0$ we have $y_t = \varepsilon_t\sqrt{\alpha_0}$ and $\text{Var}(y_t) = \alpha_0$ and the model reduces to white noise or a normal distribution under the assumption of normal errors. If $\alpha_1 > 0$ we see from (2.13) that both the second and fourth moments will exceed that of the normal distribution (if $\alpha_1 = 0$). Thus the variance as well as the kurtosis of the ARCH(1) model exceeds that of the normal distribution. It follows that the ARCH(1) model is leptokurtic with *fat tails* as depicted in Figure 2.2.

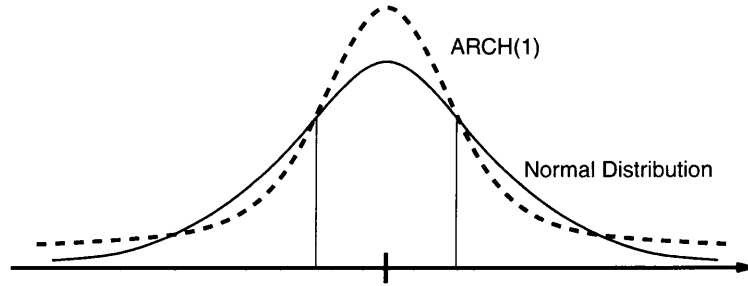


FIGURE 2.2. The Normal and ARCH(1) distributions

2.3. The ARCH(p) Model. The ARCH(1) model's memory is only one period but we often require a lengthy lag structure to adequately model volatility clustering of financial time series. It is easy to expand this memory to p lags as in the definition (2.2).

THEOREM 2.11. *The mean and all other odd moments of the ARCH(p) model are zero.*

PROOF.

$$E[y^{2k+1}|\Psi_{t-1}] = h_t^{\frac{2k+1}{2}} E[\varepsilon_t]^{2k+1} = 0.$$

As this does not depend on t it also gives the unconditional moments. $\square$

THEOREM 2.12. *The p th-order ARCH(p) model has finite variance*

$$\gamma_0 = \frac{\alpha_0}{1 - \alpha_1 - \alpha_2 - \cdots - \alpha_p}. \quad (2.15)$$

if, and only if, the associated characteristic equation has all roots outside the unit circle.

PROOF. The conditional mean of y_t^2 is

$$\begin{aligned} \gamma_0|\Psi_{t-1} &= E[y_t^2|\Psi_{t-1}] \\ &= E[\varepsilon^2 h_t|\Psi_{t-1}] \end{aligned}$$

and, since ε_t is independent of h_t

$$\begin{aligned} &= E[\varepsilon^2|\Psi_{t-1}] E[h_t|\Psi_{t-1}] \\ &= E[\alpha_0 + \alpha_1 y_{t-1}^2 + \alpha_2 y_{t-2}^2 + \cdots + \alpha_p y_{t-p}^2|\Psi_{t-1}]. \end{aligned}$$

If we define

$$\mathbf{w}_t = \begin{pmatrix} y_t^2 \\ y_{t-1}^2 \\ y_{t-2}^2 \\ \vdots \\ y_{t-p}^2 \end{pmatrix}$$

we can write

$$E[\mathbf{w}_t | \Psi_{t-1}] = \mathbf{b} + \mathbf{A} \mathbf{w}_{t-1}$$

with

$$\mathbf{b} = \begin{pmatrix} \alpha_0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

and

$$\mathbf{A} = \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_{p-1} & \alpha_p \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

This matrix equation in y_t^2 is similar to the difference equation (1.54) in y_t . After applying an infinite back substitution we obtain the unconditional expectation as

$$\begin{aligned} \lim_{k \rightarrow \infty} E[y_t^2 | \Psi_{t-k}] &= \lim_{k \rightarrow \infty} (\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^{k-1}) \mathbf{b} + \mathbf{A}^k \mathbf{w}_{t-k} \\ &= (\mathbf{I} - \mathbf{A})^{-1} \mathbf{b} \end{aligned}$$

only if the eigenvalues of $\mathbf{A}$ are within the unit circle by Theorem 1.55. This is equivalent to requiring that the roots of the characteristic equation are

outside the unit circle by Corollary 1.59. The first element of $(\mathcal{I} - \mathbf{A})^{-1}$ from Theorem 1.56 is

$$\frac{1}{1 - \alpha_1 - \alpha_2 - \cdots - \alpha_p}$$

and if we multiply this with $\mathbf{b}$ we get the required result. $\square$

From this proof it is clear that there is a close relationship between the y_t^2 of the ARCH(p) model and the y_t of the difference equation or the AR(p) model. We will explore this relationship further once we have further generalized the ARCH(p) model.

3. GARCH Models

Earlier we combined AR(p) and MA(q) models in order to obtain a more general and often more parsimonious ARMA class of models. In a like manner we generalize ARCH models after Bollerslev (1986) to obtain *generalized autoregressive conditional heteroscedasticity* or GARCH time series models.

DEFINITION 2.13. The GARCH(p, q) time series model is defined as

$$y_t = \varepsilon_t \sqrt{h_t}$$

with ε_t white noise and the added restriction that

$$\text{Var}(\varepsilon_t) = 1.$$

Further we define h_t more generally as

$$\begin{aligned} h_t &= \alpha_0 + \sum_{i=1}^q \alpha_i y_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i} \\ &= \alpha_0 + \alpha(B) y_t^2 + \beta(B) h_t \end{aligned}$$

with B the backward shift operator and $\alpha_0 > 0$ and $\alpha_i, \beta_i \geq 0$.

Usually it is assumed that $\varepsilon_t \sim N(0, 1)$ so that we can write

$$y_t | \Psi_{t-1} \sim N(0, h_t).$$

We now also include previous observations of h_t in the model and use these to predict future h_t 's.

3.1. Stationarity Conditions.

THEOREM 2.14. *A GARCH(p, q) model is stationary with variance*

$$\gamma_0 = \frac{\alpha_0}{1 - \alpha(1) - \beta(1)} \quad (2.16)$$

if and only if $\alpha(1) + \beta(1) < 1$.

PROOF. The mean of the GARCH(p, q) model is zero and we only have to show that the variance is finite. As in Bollerslev (1986:323) we substitute $y_t = \varepsilon_t \sqrt{h_t}$ in h_t

$$\begin{aligned} h_t &= \alpha_0 + \sum_{i=1}^q \alpha_i y_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i} \\ &= \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 h_{t-i} + \sum_{i=1}^p \beta_i h_{t-i} \\ &= \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 \left(\alpha_0 + \sum_{j=1}^q \alpha_j \varepsilon_{t-i-j}^2 h_{t-i-j} + \sum_{j=1}^p \beta_j h_{t-i-j} \right) \\ &\quad + \sum_{i=1}^p \beta_i \left(\alpha_0 + \sum_{j=1}^q \alpha_j \varepsilon_{t-i-j}^2 h_{t-i-j} + \sum_{j=1}^p \beta_j h_{t-i-j} \right) \\ &\quad \vdots \\ &= \alpha_0 \sum_{k=0}^{\infty} M(t, k) \end{aligned}$$

where $M(t, k)$ involves all the terms of the form

$$\prod_{i=1}^q \alpha_i^{a_i} \prod_{j=1}^p \beta_j^{b_j} \prod_{l=1}^n \varepsilon_{t-S_l}^2$$

for

$$n = \sum_{i=1}^q a_i$$

and a_i and b_j such that

$$\sum_{i=1}^q a_i + \sum_{j=1}^p b_j = k$$

and

$$1 \leq S_1 < S_2 < \cdots < S_n \leq \max[kq, (k-1)q + p].$$

Thus we have

$$\begin{aligned} M(t, 0) &= 1, \\ M(t, 1) &= \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i, \\ M(t, 2) &= \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 \left(\sum_{j=1}^q \alpha_j \varepsilon_{t-i-j}^2 + \sum_{j=1}^p \beta_j \right) \\ &\quad + \sum_{i=1}^p \beta_i \left(\sum_{j=1}^q \alpha_j \varepsilon_{t-i-j}^2 + \sum_{j=1}^p \beta_j \right) \\ &= \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 M(t-i, 1) + \sum_{i=1}^p \beta_i M(t-i, 1) \end{aligned}$$

and in general

$$M(t, k+1) = \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 M(t-i, k) + \sum_{i=1}^p \beta_i M(t-i, k).$$

Since $M(t, k)$ is a linear combination of the i.i.d. ε_t^2 its moments are not dependent on t . In particular

$$\mathbb{E}[M(t, k)] = \mathbb{E}[M(s, k)]$$

and, since $E[\varepsilon_t^2] = 1$ we obtain the recursive equation

$$\begin{aligned}
 E[M(t, k+1)] &= E \left[\sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 M(t-i, k) + \sum_{i=1}^p \beta_i M(t-i, k) \right] \\
 &= \left(\sum_{i=1}^q \alpha_i + \sum_{i=1}^p \beta_i \right) E[M(t, k)] \\
 &= \left(\sum_{i=1}^q \alpha_i + \sum_{i=1}^p \beta_i \right)^2 E[M(t, k-1)] \\
 &\vdots \\
 &= \left(\sum_{i=1}^q \alpha_i + \sum_{i=1}^p \beta_i \right)^{k+1} E[M(t, 0)] \\
 &= \left(\sum_{i=1}^q \alpha_i + \sum_{i=1}^p \beta_i \right)^{k+1} \\
 &= (\alpha(1) + \beta(1))^{k+1}.
 \end{aligned}$$

From this it follows that

$$\begin{aligned}
 E[y_t^2] &= E[\varepsilon_t^2 h_t] \\
 &= E \left[\alpha_0 \sum_{k=0}^{\infty} M(t, k) \right] \\
 &= \alpha_0 \sum_{k=0}^{\infty} E[M(t, k)] \\
 &= \frac{\alpha_0}{1 - \alpha(1) - \beta(1)}
 \end{aligned}$$

if, and only if

$$\alpha(1) + \beta(1) = \sum_{i=1}^q \alpha_i + \sum_{i=1}^p \beta_i < 1.$$

□

If all the α_i and β_i are zero then y_t will simply be white noise with variance α_0 . If this is not the case it follows from (2.16) that the GARCH(p, q) model's variance exceeds that of the white noise e.g. a normal distribution.

3.2. The GARCH(1, 1) Model.

DEFINITION 2.15. The simplest GARCH model is obtained by letting $p = q = 1$ in Definition 2.13 yielding

$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1} \quad (2.17)$$

with $\alpha_0 > 0$ and $\alpha_1, \beta_1 \geq 0$.

We require $\alpha_1 + \beta_1 < 1$ for stationarity which also implies $\beta_1 < 1$. Through back substitution we can write (2.17) as an infinite-order ARCH(∞) model

$$\begin{aligned} h_t &= \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 (\alpha_0 + \alpha_1 y_{t-2}^2 + \beta_1 h_{t-2}) \\ &= \dots \\ &= \lim_{k \rightarrow \infty} \left(\alpha_0 \sum_{i=0}^k \beta_1^i + \alpha_1 \sum_{i=0}^k \beta_1^i y_{t-i-1}^2 \right) \\ &= \frac{\alpha_0}{1 - \beta_1} + \frac{\alpha_1}{1 - \beta_1} y_t^2. \end{aligned}$$

Since $\beta_1 < 1$ the weights of y_{t-i}^2 are decreasing in this infinite sum and we can approximate the GARCH(1, 1) process to any degree of accuracy using an ARCH(q) process of sufficiently high order.

3.2.1. *Moments of the GARCH(1, 1) Model.* The conditions for the existence of higher order moments for GARCH(1, 1) models as well as their general form were derived by Teräsvirta (1996). Although Bollerslev (1986) derived the same results they depended on the assumption of normal errors. Teräsvirta did away with this requirement through the following theorem.

THEOREM 2.16. Let $v_{2k} = E[\varepsilon_t^{2k}] < \infty$ and $\mu_{2k} = E[y_t^{2k}]$. If we define

$$\tau_k = E[\beta_1 + \alpha_1 \varepsilon_t^2]^k$$

and assume $\tau_k < 1$ the $2k$ th moment of the GARCH(1, 1) model has the form

$$\mu_{2k} = \frac{v_{2k}}{1 - \tau_k} \sum_{i=1}^k \binom{k}{i} \alpha_0^i \tau_{k-i} \frac{\mu_{2(k-i)}}{v_{2(k-i)}}. \quad (2.18)$$

PROOF. Note that we can write the GARCH(1, 1) model as

$$\begin{aligned}
 h_t &= \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1} \\
 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 h_{t-1} + \beta_1 h_{t-1} \\
 &= \alpha_0 + c_{t-1} h_{t-1}
 \end{aligned} \tag{2.19}$$

with $c_t = \alpha_1 \varepsilon_t^2 + \beta_1$ so that

$$\begin{aligned}
 h_t^k &= (\alpha_0 + c_{t-1} h_{t-1})^k \\
 &= c_{t-1}^k h_{t-1}^k + \sum_{i=1}^k \binom{k}{i} \alpha_0^i (c_{t-1} h_{t-1})^{k-i}.
 \end{aligned}$$

If we repeatedly substitute (2.19) into the first term we get

$$\begin{aligned}
 h_t^k &= c_{t-1}^k (\alpha_0 + c_{t-2} h_{t-2})^k + \sum_{i=1}^k \binom{k}{i} \alpha_0^i (c_{t-1} h_{t-1})^{k-i} \\
 &= c_{t-1}^k c_{t-2}^k h_{t-2}^k + \sum_{i=1}^k \binom{k}{i} \alpha_0^i (c_{t-1}^{k-i} h_{t-1}^{k-i} + c_{t-1}^k c_{t-2}^{k-i} h_{t-2}^{k-i}) \\
 &= c_{t-1}^k c_{t-2}^k c_{t-3}^k h_{t-3}^k + \sum_{i=1}^k \binom{k}{i} \alpha_0^i (c_{t-1}^{k-i} h_{t-1}^{k-i} + c_{t-1}^k c_{t-2}^{k-i} h_{t-2}^{k-i} \\
 &\quad + c_{t-1}^k c_{t-2}^k c_{t-3}^{k-i} h_{t-3}^{k-i}) \\
 &\quad \vdots \\
 &= \left(\prod_{i=1}^m c_{t-i}^k \right) h_{t-m}^k + \sum_{i=1}^k \binom{k}{i} \alpha_0^i \left[\sum_{j=1}^m \left(\prod_{l=1}^{j-1} c_{t-l}^k \cdot c_{t-j}^{k-i} h_{t-j}^{k-i} \right) \right].
 \end{aligned}$$

Now c_t is independent of h_t and also of time so that $\tau_k = E[c_t^k] = E[c_{t-i}^k]$.

Thus we can write

$$\begin{aligned}
 E[h_t^k] &= \left(\prod_{i=1}^m E[c_{t-i}^k] \right) E[h_{t-m}^k] \\
 &\quad + \sum_{i=1}^k \binom{k}{i} \alpha_0^i \left[\sum_{j=1}^m \left(\prod_{l=1}^{j-1} E[c_{t-l}^k] \cdot E[c_{t-j}^{k-i}] \cdot E[h_{t-j}^{k-i}] \right) \right] \\
 &= \tau_k^m E[h_{t-m}^k] + \sum_{i=i}^k \binom{k}{i} \alpha_0^i \left[\sum_{j=1}^m \tau_k^{j-1} \tau_{k-i} E[h_{t-j}^{k-i}] \right]
 \end{aligned}$$

where we assume that the process starts at some finite value infinitely many periods ago so. If we let $m \rightarrow \infty$ we have

$$\begin{aligned}\mu_{2k} &= E[\varepsilon_t^{2k}] E[h_t^k] \\ &= E[\varepsilon_t^{2k}] \left(\sum_{i=1}^k \binom{k}{i} \alpha_0^i \frac{\tau_{k-i}}{1 - \tau_k} E[h_t^{k-i}] \right) \\ &= \frac{v_{2k}}{1 - \tau_k} \sum_{i=1}^k \binom{k}{i} \alpha_0^i \tau_{k-i} \frac{\mu_{2(k-i)}}{v_{2(k-i)}}.\end{aligned}$$

□

This useful theorem allows us to calculate the fourth moment as well as kurtosis of a GARCH(1,1) model where we only assume that ε_t is white noise with finite second and fourth moments and finite kurtosis.

COROLLARY 2.17. *If $\tau_2 = E[\alpha_1 \varepsilon_t^2 + \beta_1]^2 < 1$ and $v_4 = E[\varepsilon_t^4] < \infty$ the fourth moment of a GARCH(1,1) model is given by*

$$\mu_4 = E[y_t^4] = \frac{\alpha_0 v_4 (1 + \tau_1)}{(1 - \tau_2)(1 - \tau_1)}$$

and the kurtosis by

$$\kappa = \frac{\mu_4}{\mu_2^2} = \kappa_\varepsilon \frac{1 - \tau_1^2}{1 - \tau_2}.$$

COROLLARY 2.18. *If $\varepsilon_t \sim N(0,1)$ then*

$$\tau_2 = E[\alpha_1 \varepsilon_t^2 + \beta_1]^2 = (\alpha_1 + \beta_1)^2 + 2\alpha_1^2$$

and if $\tau_2 < 1$ the fourth moment of the GARCH(1,1) model is

$$\mu_4 = \frac{3\alpha_0^2(1 + \alpha_1 + \beta_1)}{(1 - \alpha_1 - \beta_1)(1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2)}$$

and the kurtosis

$$\kappa = \frac{3(1 - (\alpha_1 + \beta_1)^2)}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2}.$$

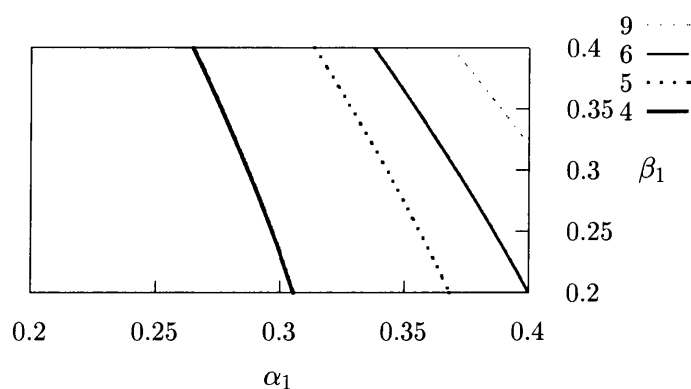


FIGURE 2.3. GARCH(1, 1) Kurtosis

This kurtosis will be 3 – that of a normal distribution – if both α_1 and β_1 are zero. If not both these parameters are zero the GARCH(1, 1) model will have a kurtosis larger than that of a normal distribution and be leptokurtic. The contours of κ are displayed in Figure 2.3 on the range $[0.2; 0.4]$ for both α_1 and β_1 .

3.3. The General GARCH(p, q) Model.

THEOREM 2.19. *If y_t follows a GARCH(p, q) model then y_t^2 follows an ARMA(r, p) model with $r = \max(p, q)$.*

PROOF. First of all, note that $E[y_t^2] = h_t$ so that the forecast error for y_t^2 – the forecast error of the GARCH model – can be defined as

$$w_t = y_t^2 - h_t \quad (2.20)$$

with this error w_t white noise. Now we can write

$$\begin{aligned}
 h_t &= \alpha_0 + \alpha_1 y_{t-1}^2 + \alpha_2 y_{t-2}^2 + \cdots + \alpha_q y_{t-q}^2 \\
 &\quad + \beta_1 h_{t-1} + \beta_2 h_{t-2} + \cdots + \beta_p h_{t-p} \\
 h_t + y_t^2 &= \alpha_0 + \alpha_1 y_{t-1}^2 + \alpha_2 y_{t-2}^2 + \cdots + \alpha_q y_{t-q}^2 \\
 &\quad - \beta_1 (y_{t-1}^2 - h_{t-1}) - \beta_1 (y_{t-2}^2 - h_{t-2}) - \cdots - \beta_p (y_{t-p}^2 - h_{t-p}) \\
 &\quad + \beta_1 y_{t-1}^2 + \beta_2 y_{t-2}^2 + \cdots + \beta_p y_{t-p}^2 + y_t^2 \\
 y_t^2 &= \alpha_0 + (\alpha_1 + \beta_1) y_{t-1}^2 + (\alpha_2 + \beta_2) y_{t-2}^2 + \cdots + (\alpha_r + \beta_r) y_{t-r}^2 \\
 &\quad + w_t - \beta_1 w_{t-1} - \beta_2 w_{t-2} - \cdots - \beta_p w_{t-p}
 \end{aligned} \tag{2.21}$$

□

This result allows us to e.g. verify Theorem 2.14 by noting that the mean for the ARMA process is

$$\mu_{\text{ARMA}} = \frac{1}{1 - \sum_{i=1}^r (\alpha_i + \beta_i)}.$$

Given the non negativity conditions and the fact that $E[\varepsilon_t^2] = \mu_{\text{ARMA}} > 0$ this is finite only if $\sum_{i=1}^r \alpha_i + \beta_i < 1$ which is the theorem.

Another useful application of this theorem is to calculate the autocovariance and autocorrelation functions of y_t^2 . From Theorem 1.112 it follows that the first order autocorrelation of y_t^2 is

$$\rho_1 = \frac{(1 - \theta_1 \phi_1)(\phi_1 - \theta_1)}{1 - 2\theta_1 \phi_1 + \theta_1^2}$$

and with $\phi_1 = \alpha_1 + \beta_1$ and $\theta_1 = \beta_1$ we have

$$= \frac{\alpha_1(1 - \alpha_1\beta_1 - \beta_1^2)}{1 - 2\alpha_1\beta_1 - \beta_1^2}.$$

If we use the result of Corollary 2.18 to, as noted by Teräsvirta (1996:9) expand

$$\begin{aligned}
 \frac{\beta_1(1 - 3\kappa^{-1})}{3(1 - \kappa^{-1})} + \alpha_1 &= \beta_1 \frac{1 - \frac{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2}{1 - (\alpha_1 + \beta_1)^2}}{3 - \frac{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2}{1 - (\alpha_1 + \beta_1)^2}} + \alpha_1 \\
 &= \frac{\alpha_1^2 \beta_1}{1 - (\alpha_1 + \beta_1)^2 + \alpha_1^2} + \alpha_1 \\
 &= \frac{\alpha_1(\alpha_1 \beta_1 + 1 - 2\alpha_1 \beta_1 - \beta_1^2)}{1 - 2\alpha_1 \beta_1 - \beta_1^2} \\
 &= \frac{\alpha_1(1 - \alpha_1 \beta_1 - \beta_1^2)}{1 - 2\alpha_1 \beta_1 - \beta_1^2} \\
 &= \rho_1.
 \end{aligned}$$

We find that with a high kurtosis this ratio becomes

$$\rho_1 \approx \alpha_1 + \frac{\beta_1}{3}.$$

Since we require a high value of α_1 in order to obtain a high kurtosis this implies that the first order autocorrelation of y_t^2 will also be high. In fact, both the kurtosis and first order autocorrelation are monotonically increasing functions of α_1 . It is well known that financial time series are usually leptokurtic with fairly high kurtosis but it is also known that there normally is low but persistent first order autocorrelation in the squared observations. As noted by Teräsvirta (1996:1) this phenomena was already pointed out by Granger and Andersen in 1978¹.

The GARCH(1, 1) model does not simultaneously allow for high kurtosis and low first order autocorrelation in the squares. In practice better results are often obtained using leptokurtic errors e.g. the Student's t distribution rather than normal errors.

3.4. Other GARCH Models. The popularity of GARCH models also gave rise to a number of alternative methods for determining h_t being

¹This can also be noted clearly in the South African market as in Figure 3.7 in the next chapter.

proposed. The two most widely used alternative models are Schwert's Autoregressive Standard Deviation model

$$h_t = \left[\alpha_0 + \sum_{i=1}^p \alpha_i |y_{t-i}| \right]^2$$

and Nelson's exponential GARCH or EGARCH model

$$\ln h_t = \omega + \beta \ln h_{t-1} + \gamma \frac{y_{t-1}}{\sqrt{h_{t-1}}} + \alpha \left[\frac{|y_{t-1}|}{\sqrt{h_{t-1}}} - \sqrt{\frac{2}{\pi}} \right].$$

Other well known models are

- Nonlinear ARCH or NARCH

$$h_t = \omega + \alpha |y_{t-1}|^\gamma + \beta h_{t-1}$$

- Multiplicative ARCH or MARCH

$$\ln h_t = \omega + \sum_{i=1}^p \alpha_i \ln y_{t-1}^2$$

- GJR (Glosten, Jagannathan and Runkle)

$$h_t = \omega + \beta h_{t-1} + \alpha y_{t-1}^2 + \gamma S_{t-1}^- y_{t-1}^2$$

with

$$S_t^- = \begin{cases} 1 & y_t < 0 \\ 0 & y_t \geq 0. \end{cases}$$

- Asymmetric GARCH or AGARCH

$$h_t = \omega + \alpha (\varepsilon_{t-1} + \gamma)^2 + \beta h_{t-1}$$

- Nonlinear Asymmetric GARCH or NAGARCH

$$h_t = \omega + \beta h_{t-1} + \alpha \left(y_{t-1} + \gamma \sqrt{h_{t-1}} \right)^2$$

- VGARCH

$$h_t = \beta h_{t-1} + \alpha \left(\frac{y_{t-1}}{\sqrt{h_{t-1}}} + \gamma \right)^2$$

Many of these alternatives aim at introducing asymmetry into the symmetric GARCH model in order to address the observed phenomena that a negative y_t often leads to a higher expected variance than a positive y_t of similar magnitude. For more of these models as well as references to the ones listed here, see Engle & Ng (1993), Engle (1993) and Teräsvirta (1997) who also discusses the properties of some of these models.

3.5. The News Impact Curve. If the GARCH type models developed in this chapter are applied to a time series of the returns of some financial instrument some interesting and important observations can be made. Employing this approach Engle and Ng (1993) defined the *news impact curve* and used this to measure how new and unexpected information is incorporated into volatility estimates and contributes to volatility increases.

By defining x_t as the rate of return from an investment from time $t - 1$ to t with Ψ_t the information set it is plausible to argue that investors, in making investment decision at time $t - 1$, employ all available information Ψ_{t-1} . They pay attention specifically to the conditional expected value of x_t namely $\mu_t = E[x_t|\Psi_{t-1}]$ as well as the conditional variance of x_t or $h_t = \text{Var}(x_t|\Psi_{t-1})$. We now define a new time series – the *unexpected return* at time t – as

$$y_t = x_t - \mu_t.$$

Engle and Ng then treat y_t as a collective measure of news arriving at time t . An unexpectedly high return $y_t > 0$ indicates the arrival of good news and similarly $y_t < 0$ indicates the arrival of bad news. The magnitude of y_t reveals the ‘significance’ of the news in the sense that a large value of $|y_t|$ hints at the arrival of ‘big’ news.

For the regular ARCH(p) model we have

$$h_t = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i}^2$$

and the effect of the arrival of new news in period $t - i$, $i \leq p$ on h_t is controlled by the parameter α_i . The older the news is the less the impact we would expect it to have on the current volatility of returns and thus we

expect that $\alpha_i < \alpha_j$ for $i > j$. The impact of news that arrived more than p periods ago is considered negligible. The GARCH(p, q) model on the other hand have

$$h_t = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i}^2 + \sum_{i=1}^q \beta_i h_{t-i}$$

which may also be expressed as an infinite sum of past ε_t or as an ARCH(∞) model. Here the impact of news in the distant past on current volatility is infinite in extent but still decreasing in magnitude.

If we now fix all information in Ψ_{t-1} but allow ε_{t-1} to vary we can determine and plot its impact on h_t . The plot of this relationship is defined as the *news impact curve* (NIC). Both the ARCH and GARCH models produce a symmetrical NIC. In practice the NIC is usually negatively asymmetric. This occurs when the arrival of unexpected bad news has a larger impact on predictable volatility than the arrival of unexpected good news of a similar magnitude. In short, the market reacts differently to good and bad news, even if the good news is just as good as what the bad news is bad. The market becomes more volatile due to the arrival of bad news than when good news arrives. This corresponds to the findings of Black (1976) and confirmed by many others.

The NIC of many of the alternative models presented in the previous section is not symmetric. Engle and Ng (1993) found that the best model to use for Japanese stock return data is the asymmetric GJR model. They also found that the EGARCH model captured the asymmetry but noted that the variability of the conditional variance implied by this model is too high.

We can construct the NIC in two ways. The conditional NIC will make use of the most recent information available in Ψ_{t-1} and as such the value of e.g. the intercept will be dictated by recent observed values for y_t and h_t . On the other hand the unconditional NIC will have as its intercept the *long run* or unconditional volatility e.g.

$$\gamma_0 = \frac{\alpha_0}{1 - \alpha_1 - \alpha_2 - \cdots - \alpha_p - \beta_1 - \beta_2 - \cdots - \beta_q}$$

from (2.16) if we use a GARCH(p, q) model. Three different, arbitrary NICs for different models are displayed in Figure 2.4.

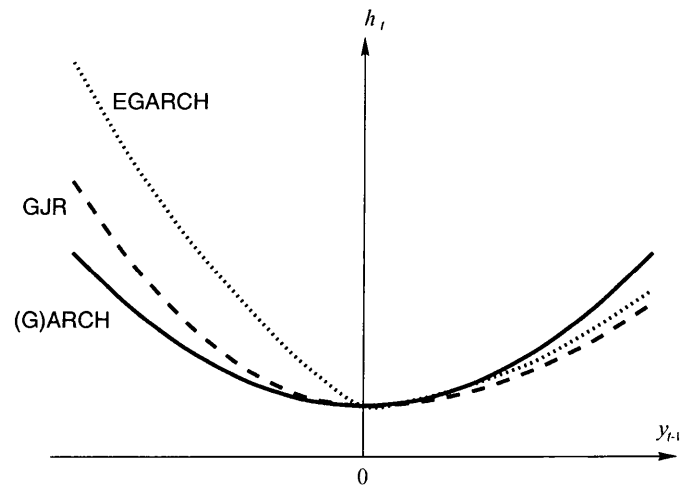


FIGURE 2.4. Arbitrary News Impact Curve for Various Models

CHAPTER 3

The South African News Impact Curve

1. Introduction

Here we will utilise the techniques developed earlier to estimate and evaluate the South African news impact curve (NIC) using various alternative ARCH and GARCH representations. The four major financial markets in the local economy are

- *Money Market* This market represents ‘safe’ cash investments. The risk in it is associated with movements in the BA rate and we will use the *negative* of the daily change in the BA rate to estimate the NIC.
- *Bond Market* This market is also interest rate driven and we will use the *negative* of changes in the benchmark R150 government stock’s daily closing yield to estimate the NIC.
- *Share Market* This market is driven by share price movements and we will use daily percentage changes in the JSE Overall Index.
- *Forex Market* This market is driven by currency price movements and here we will use the *negative* of daily percentage changes in the ZAR/USD exchange rate.

We are specifically interested in the asymmetric shape of the NIC and thus we want to use time series in such a way that a negative value represents ‘bad’ news for the local markets. Higher interest rates or yields in the interest rate markets are indicative of a lower capital value. Similarly a rise in the quoted ZAR/USD exchange rate implies a depreciation in the value of the Rand relative to the Dollar. These are the reasons why we use the negative of the money, bond and forex market indices as indicated above.

For brevity we will refer to these four markets as MM, BM, SM and FM respectively and will for example say that the standard deviation of SM was 0.815% during 1989. We will use a sample of 2,500 observations from

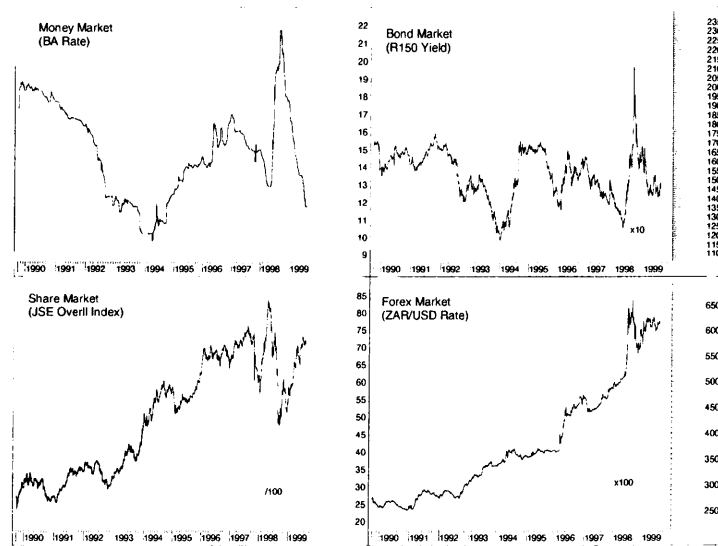


FIGURE 3.1. Actual South African Market Indices

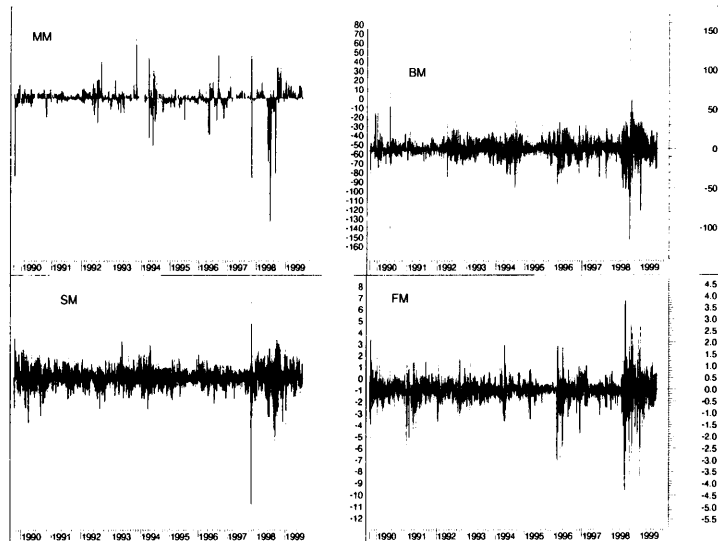


FIGURE 3.2. Transformed Market Indices

19891003 to 19990813. The various actual indices are displayed in Figure 3.1 and the transformed series we will utilise in Figure 3.2. Keep in mind that an upward movement in the transformed series always indicates 'good' news for South Africa and vice versa. All the transformed series have been multiplied by 100.

Statistic	MM	BM	SM	FM
Average	0.225%	0.079%	0.044%	-0.034%
Standard Deviation	9.427%	13.135%	1.057%	0.546%
Variance	89.717%	172.527%	1.118%	0.298%
Skewness	-4.853%	-0.362%	-1.231%	-0.936%
Kurtosis	73.580%	16.726%	13.746%	13.166%
Ljung-Box	386.919	144.401	118.008	139.201

TABLE 3.1. Descriptive Statistics for MM, BM, SM and FM

2. Fitting ARMA Models

2.1. The SAC and SPAC. We start off by fitting ARMA models to MM, BM, SM and FM. As part of the ARMA identification process we need to look at the SAC and SPAC of each of these markets, as in Figure 3.3 and 3.4. It seems appropriate to use some ARMA model for MM but it is more difficult to determine tentative models for BM, SM and FM as *both* the SAC and SPAC cuts off. If the SAC cuts off some practitioners consider the behaviour of the SPAC irrelevant and only fit an appropriate MA model. Bowerman and O'Connell (1987:73) suggest that both AR and MA models be considered. They mention that if both the SAC and SPAC cuts off we should try to determine which cuts off more abruptly. If both cuts off equally abruptly then we should fit both AR and MA models and select the best of these.

2.2. Likelihood Estimation of ARMA Model Parameters. The likelihood function is the product of the probabilities of each outcome realised in the sample and thus the probability of realising the sample

$$L = P(y_1, y_2, \dots, y_n) = f_1(y_1) \cdot f_2(y_2) \cdot \dots \cdot f_n(y_n).$$

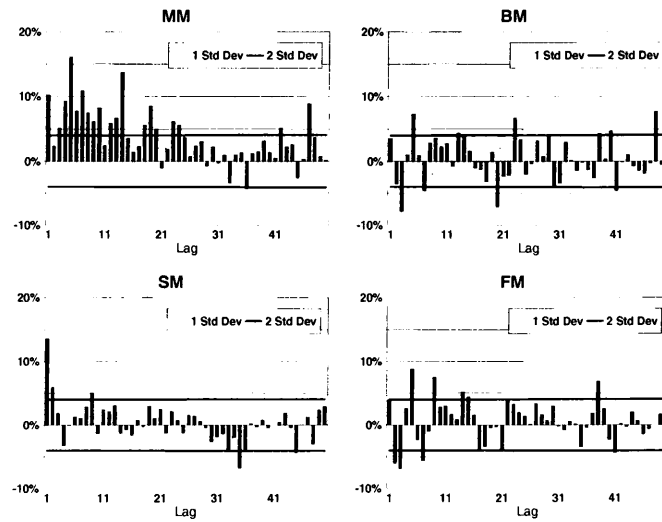


FIGURE 3.3. Sample Autocorrelation Function

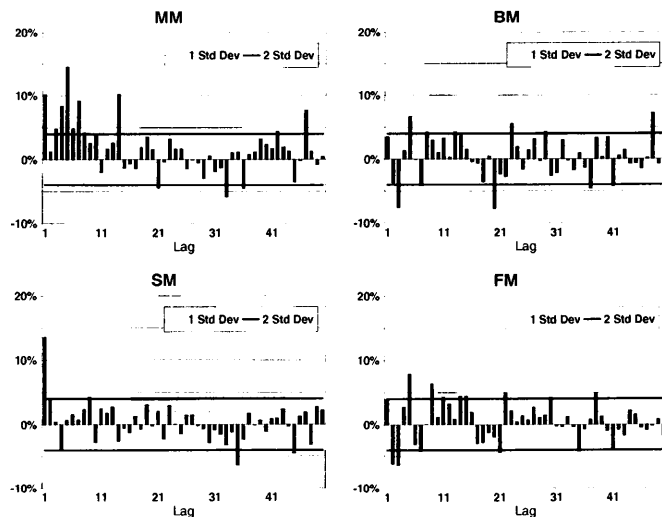


FIGURE 3.4. Sample Partial Autocorrelation Function

We will maximise the (log) likelihood function of the ARMA model over the parameter space in order to estimate the model parameters

$$\theta = (\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q)'$$

Thus we maximise the probability of the given sample being realised. If we assume that the errors are normal the likelihood function is

$$L(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n; \boldsymbol{\theta}) = \left(\frac{1}{\sqrt{2\pi\sigma_\varepsilon}} \right)^n e^{-\sum_{i=1}^n \frac{\varepsilon_i^2}{2\sigma_\varepsilon^2}}.$$

We need to start up the ARMA(p, q) equation using p previous y_t 's and q previous ε_t 's. There is a number of ways in which to address this problem – we will start estimating from observation $p + 1$ and use the first p actual y_t 's to start up the equation. The previous q errors will be set equal to the unconditional mean which is zero. If we define

$$S = \sum_{i=p+1}^n \varepsilon_i^2$$

as the sum of the $n' = n - p$ squared errors the log likelihood function is

$$l = \ln L = -\frac{1}{2}n' \ln 2\pi - \frac{1}{2}n' \ln \sigma_\varepsilon^2 - \frac{S}{2\sigma_\varepsilon^2} = -c_0 - \frac{S}{2\sigma_\varepsilon^2}. \quad (3.1)$$

Maximising the log likelihood function is thus similar to minimising the sum of squared errors so that the least squares estimates are also the maximum likelihood estimates.

A simple Nelder-Mead algorithm¹ (Mathews 1987:295) was used to optimise the various alternative models. Since this algorithm does not use derivatives it proved to be very robust, especially later on when the GARCH type models was estimated. The optimisation was however quite slow, both in terms of the time and number of iterations it needed to converge.

2.3. Statistical Inference with Maximum Likelihood Estimation. In order to comment on the significance of parameter estimates we need to evaluate the information matrix. This matrix is

$$\mathbf{I} = - \lim_{n \rightarrow \infty} E \left[\frac{1}{n'} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] \quad (3.2)$$

with l the log likelihood function (3.1). The asymptotic distribution of maximum likelihood estimators (Campbell, Lo & MacKinlay 1997:538) is

$$\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}_0, \frac{1}{n'} \mathbf{I}^{-1}) \quad (3.3)$$

¹The algorithm was initially implemented using the perl programming language but after a few initial tests it was moved to the Yorick programming language which improved the execution speed significantly.

with $\mathbf{I}$ positive definite in the region of the optimum $\boldsymbol{\theta}_0$.

We will make use of (3.3) to calculate the covariance matrix of the estimates. For ARMA models with S the sum of squared errors and

$$\mathcal{S} = \frac{\partial^2 S}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \quad (3.4)$$

we can approximate this covariance matrix using (3.1) through

$$\begin{aligned} \text{Var}(\boldsymbol{\theta}) &\approx -\frac{1}{n'} \mathbf{I}^{-1} \\ &= 2\sigma_\varepsilon^2 \mathcal{S}^{-1}. \end{aligned}$$

2.4. The Ljung-Box Statistic. In order to discriminate between the various models we need an arsenal of diagnostic statistics, one of which is the *Ljung-Box statistic*.

DEFINITION 3.1. The Ljung-Box statistic is defined as

$$Q^* = n'(n' + 2) \sum_{i=1}^K \frac{r_i^2(\varepsilon)}{n' - i} \quad (3.5)$$

where $n' = n - d$ with d the order of differencing applied and $r_i^2(\varepsilon)$ the autocorrelation in the error term.

If the model captures the interrelationships in the data the autocorrelation in the error will be low and Q^* will be small. A large Q^* thus indicates an inadequate model. We can also perform inference testing on this statistic as it follows a $\chi^2(K - k)$ distribution with k the number of model parameters (Bowerman & O'Connell 198:149).

We choose $K = 50$ arbitrarily – the same as the number of lags displayed in the SAC and SPAC. In practise it is often a good idea to evaluate this statistic for a number of different K 's reflecting the possible seasonality in the data, e.g. $K = 12, 24, \dots$ for monthly data. Since

$$E[Q^*(K_1) - Q^*(K_2)] = K_1 - K_2$$

this statistic can also be used to check different parts of the correlogram (Kendall & Ord 1990:112).

2.5. Information Criteria. The Ljung-Box statistic does not take the amount of model parameters into account. Since we want to penalise models with more parameters we need to make use of this in our diagnostic statistics. A natural approach is to compare the likelihood functions of different models. If the likelihood L_1 of a model was arrived at using k_1 parameters and L_2 using k_2 parameters it is well known that for large samples (Kendall & Ord 1990:117)

$$E\left[-2 \ln \frac{L_1}{L_2}\right] = k_2 - k_1.$$

This gave rise to a number of statistics such as Akaike's information criterion

$$AIC = -2 \ln L + 2k \quad (3.6)$$

and Schwartz' Bayesian information criterion

$$BIC = -2 \ln L + k \ln n. \quad (3.7)$$

The model with the smallest AIC or BIC is deemed best. The AIC should decrease with 2 for each additional parameter while the BIC should decrease with roughly $\ln 2,500$ or 7.8 in our case. The BIC normally penalises a model with more parameters heavier than the AIC and it also has better asymptotic properties than the AIC .

2.6. Selecting the Final ARMA Type Models. After inspecting the SAC and SPAC of the various markets (Figure 3.3 and 3.4) we proceeded to fit $AR(k)$, $MA(k)$ and $ARMA(k, k)$ models for $k = 1, 2, \dots, 10$ to MM, BM, SM and FM and calculated the Ljung-Box and BIC statistics for each of these as depicted in Figure 3.5 and 3.6.

There were some problems in obtaining adequate convergence in especially the higher order ARMA models which can be discerned from Figure 3.5 (e.g. problems with $k = 10$ for BM). The reasoning behind the final selection is given next with the salient model parameters in Table 3.2. We require that the roots of $\phi(z)$ and $\theta(z)$ be outside the unit circle for stationarity and invertibility respectively. A necessary but not sufficient check for

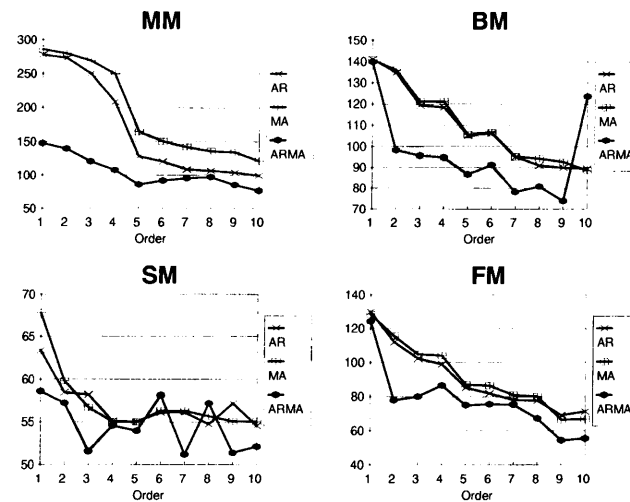
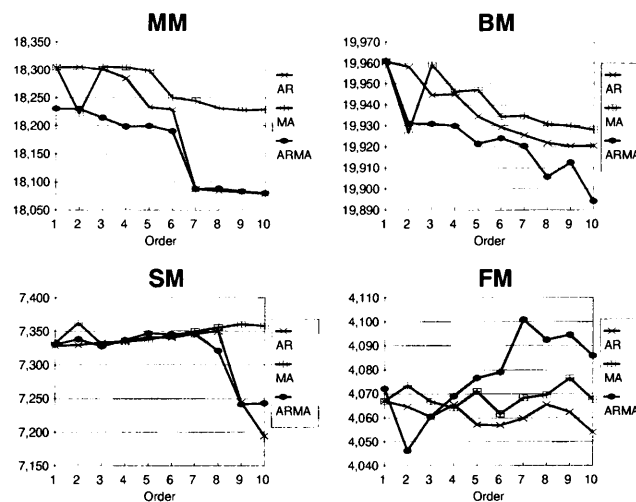


FIGURE 3.5. Ljung-Box Statistic

FIGURE 3.6. Bayesian Information Criterion (*BIC*)

both these conditions (Kendall & Ord 1990:112) is that

$$|\phi_p| < 1, \quad |\theta_q| < 1, \quad \phi(1) > 0, \quad \phi(-1) > 0, \quad \theta(1) > 0 \quad \text{and} \quad \theta(-1) > 0.$$

The first two conditions are easy to verify and the results of the last four are shown in Table 3.2. Although these conditions are not sufficient there are no obvious breaches.

- *MM* The SAC and SPAC suggested some ARMA model and this is confirmed by the Ljung-Box and *BIC* although for higher order ($k \geq 5$) models there is little difference between ARMA and AR models. The AR(5) or AR(7) models present themselves as candidates from these statistics and the AR(7) model was selected mainly because of the big drop in the *BIC* at this order.
- *BM* The SAC and SPAC do not provide clear direction here but the Ljung-Box and *BIC* strongly suggests an ARMA(2, 2) model and this was selected.
- *SM* Here the SAC and SPAC also were not decisive. The Ljung-Box and *BIC* is not clear either, apart from suggesting that a suitable AR or MA model should provide better or at least similar results to an ARMA model. Based on the Ljung-Box an MA(4) model was finally selected.
- *FM* Against the SAC and SPAC lacks clear direction but both the Ljung-Box and *BIC* suggested an ARMA(2, 2) model and this was chosen.

3. Fitting GARCH Models

We now create four new time series y_t for each of MM, BM, SM and FM. These time series are simply the errors generated by the ARMA models fitted in the previous section². Here we will fit suitable GARCH type models to these errors and then calculate the NIC from these models over various time periods.

We will firstly fit a GARCH type model for the full time period. After this we split the sample into roughly two equal portions at 19940601. We fit the same model to the pre and post 19940601 sample. The reason for choosing this date is not completely arbitrary. This year was a year of major

²It is important to note the change over in notation. In the previous section we used y_t for the transformed daily returns recorded in each market and referred to this as MM, BM, SM and FM. Then we fitted an ARMA model to each market and obtained the errors ε_t of the fit. In this section y_t and similarly MM, BM, SM and FM refers to these error terms and ε_t to the GARCH error.

	<i>MM</i>	<i>BM</i>	<i>SM</i>	<i>FM</i>
$\bar{y}_t$	0.225	0.079	0.044	-0.034
$\sum \varepsilon_i^2$	202,651.490	425,085.393	2,731.773	730.230
σ_ε^2	81.352	170.307	1.094	0.293
$l = \ln L$	-9,022.837	-9,958.722	-3,657.208	-2,009.915
Ljung-Box	109.141	98.312	55.035	77.932
<i>BIC</i>	18,100.419	19,948.735	7,345.712	4,051.122
R^2	9.6%	1.4%	2.2%	1.9%
$\phi(1)$	0.571	1.505	1.000	1.391
$\phi(-1)$	1.244	2.191	1.000	2.198
$\theta(1)$	1.000	1.504	1.181	1.355
$\theta(-1)$	1.000	2.093	0.874	2.076
	$\phi_1 = 0.070$ (0.019, 3.6)	$\phi_1 = 0.343$ (0.062, 5.5)	$\theta_1 = -0.131$ (0.020, -6.5)	$\phi_1 = 0.403$ (0.071, 5.7)
	$\phi_2 = -0.015$ (0.020, -0.8)	$\phi_2 = -0.849$ (0.052, -16.3)	$\theta_2 = -0.060$ (0.020, -3.0)	$\phi_2 = -0.795$ (0.059, -13.5)
	$\phi_3 = 0.030$ (0.020, 1.6)	$\theta_1 = 0.295$ (0.070, 4.2)	$\theta_3 = -0.023$ (0.020, -1.1)	$\theta_1 = 0.360$ (0.083, 4.3)
	$\phi_4 = 0.066$ (0.019, 3.4)	$\theta_2 = -0.798$ (0.061, -13.2)	$\theta_4 = 0.032$ (0.020, 1.6)	$\theta_2 = -0.715$ (0.066, -10.8)
	$\phi_5 = 0.144$ (0.019, 7.5)			
	$\phi_6 = 0.041$ (0.020, 2.1)			
	$\phi_7 = 0.092$ (0.020, 4.7)			

TABLE 3.2. ARMA Model Parameters

political transition for South Africa after which global participation in the local financial markets rapidly increased. It will be interesting to note how this affected the symmetry of the NIC.

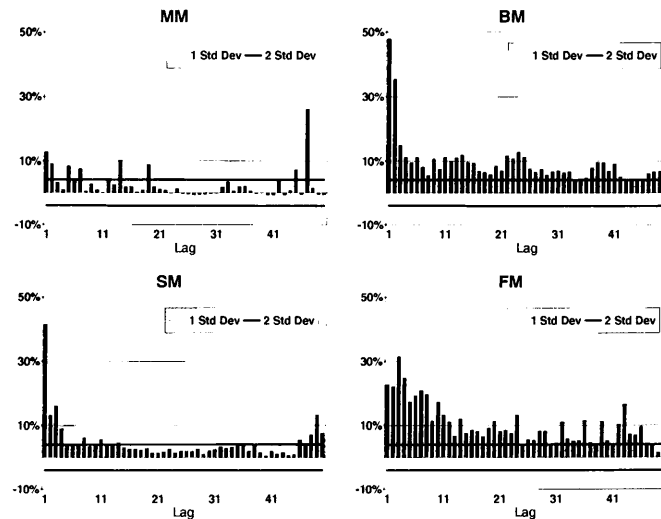


FIGURE 3.7. SAC of Squared Errors

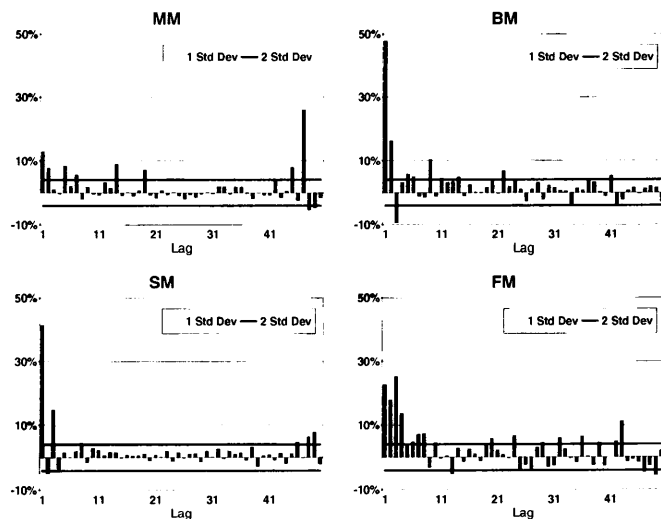


FIGURE 3.8. SPAC of Squared Errors

In order to determine the GARCH model order we follow much the same procedure as before with the ARMA models. Using Theorem 2.19 we first of all examine the SAC and SPAC of the squared errors y_t^2 – these appear in Figure 3.7 and 3.8. Tentatively a second order model will suffice for MM and a third order model for BM. SM and FM each requires a higher fourth order model.

3.1. The GARCH Maximum Likelihood Function. In order to further assist in determining the correct model to use as well as to estimate the model parameters we need to derive the likelihood function of GARCH models. If $y_t = \varepsilon_t \sqrt{h_t}$ and we assume that $\varepsilon_t \sim N(0, 1)$ we have

$$f(y_t | \Psi_{t-1}) = \frac{1}{\sqrt{2\pi h_t}} e^{-\frac{(y_t - \mu_t)^2}{2h_t}}.$$

Since $\mu_t = E[y_t] = 0$ the likelihood function becomes

$$L = \prod_{i=1}^n \frac{1}{\sqrt{2\pi h_i}} e^{-\frac{y_i^2}{2h_i}} \quad (3.8)$$

and the log likelihood

$$l = -\frac{n}{2} \ln 2\pi - \frac{1}{2} \sum_{i=1}^n \ln h_i - \frac{1}{2} \sum_{i=1}^n \frac{y_i^2}{h_i}. \quad (3.9)$$

In order to start up the system we will again make use of the first few observations and then find the parameters by maximising the log likelihood over the parameter space using the same Nelder-Mead algorithm we employed before.

3.2. Maximizing the Likelihood Function. This time the optimisation is constrained as none of the GARCH parameters are allowed to assume negative values. The Nelder-Mead algorithm does not make use of derivatives and this proved to be a handy attribute in performing the constrained optimisation.

Figure 3.8 shows that both BM and SM have significant negative partial autocorrelations of a low order that in turn may require negative GARCH parameters. This is not allowed and introduces some problems as the information matrix no longer is positive definite if we obtain an optimal point within the constraints that is not close to the real optimum.

If the real optimum lies at a point that contains negative parameter values the constrained optimisation will force those parameter values close to zero. The information matrix at this point will be numerically close to singular and not positive definite. The matrix inversion procedure used to obtain the parameter covariance matrix (3.3) is the robust singular value decomposition algorithm (Press et al. 1989:60) that allows us to invert such

ill-conditioned matrices. Whenever this problem was encountered the problematic parameter and corresponding row and column from the information matrix were removed.

3.3. The final GARCH Models and NIC. In order to assist in selecting the final GARCH models for each market we will fit $\text{GARCH}(k, k)$ models with $k = 1, \dots, 9$ in turn and calculate the Ljung-Box and BIC statistics for each market. Since this is still the identification stage we will optimise using a fairly large error and relatively few iterations. Thus the results will be approximate but they do serve the purpose of model identification.

We will fit the following eight different GARCH models identified in Section 3.4:

- Symmetric GARCH: Regular GARCH model
- Symmetric NARCH: Non linear ARCH
- Asymmetric GJR: Glosten, Jagannathan and Runkle
- Asymmetric EGARCH: Exponential GARCH
- Symmetric ASTDDEV: Autoregressive standard deviation
- Asymmetric AGARCH: Asymmetric GARCH
- Asymmetric NAGARCH: Non linear asymmetric GARCH
- Asymmetric VGARCH.

3.3.1. *MM.* For MM Figure 3.9 was constructed. Since the mean is zero the frequency distribution clearly shows a negative skew and leptokurtic distribution indicating the need for an asymmetric model. The Ljung-Box and BIC contains the eight different model types with the thick line representing the regular GARCH model.

The Ljung-Box is requesting a fourth order and the BIC a fifth order model. At these orders there is no clear distinction between the symmetric GARCH model and the rest which is somewhat disappointing. Revisiting the SAC and SPAC and looking more closely at the Ljung-Box statistic justifies a second order asymmetric model. At this order both the Ljung-Box and BIC of the AGARCH model is relatively low and this is the asymmetric

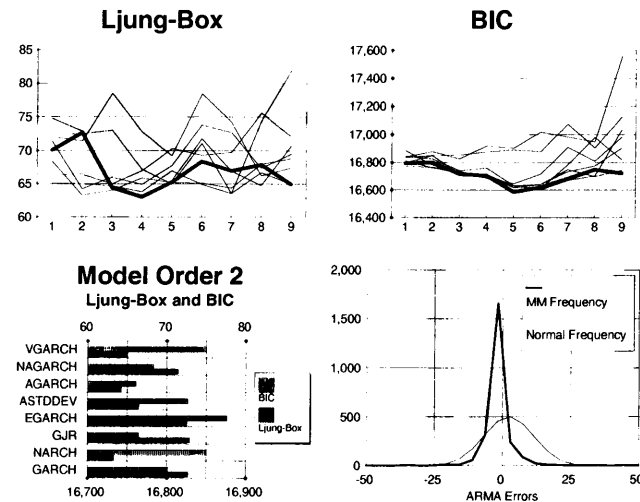


FIGURE 3.9. MM GARCH Models at Order 2

model chosen

$$h_t = \alpha_0 + \alpha_1(\varepsilon_{t-1} - \gamma)^2 + \alpha_2\varepsilon_{t-2}^2 + \beta_1h_{t-1} + \beta_2h_{t-2}.$$

Since the estimate of α_2 was close to zero the corresponding term was dropped from the model and we finally used an AGARCH(1,2) model

$$h_t = \alpha_0 + \alpha_1(\varepsilon_{t-1} - \gamma)^2 + \beta_1h_{t-1} + \beta_2h_{t-2}. \quad (3.10)$$

It is interesting to note that this time the α_0 parameter was approaching the border value of 0. Since we require such a constant parameter it was not dropped from the final equation.

The values of the parameters as well as diagnostics for the various MM models are given in Table 3.3. As a check for stationarity we also give the sum of the α_i and β_i parameters which should be less than unity. Note that this is not the case in general which is a typical problem with GARCH models in practise.

The asymmetric parameter γ is significant and there is little difference in this parameter between the two halves of the sample. There is however a big difference between this parameter in any of the two halves when compared to the full sample estimate.

	<i>Full Sample</i>		<i>Pre 19940601</i>		<i>Post 19940601</i>	
	GARCH	AGARCH	GARCH	AGARCH	GARCH	AGARCH
$\bar{y}_t^2$	81.321	81.321	46.088	46.088	120.554	120.554
$\bar{h}_t$	105.863	98.418	53.604	51.641	181.312	198.714
$\sum \varepsilon_i^2$	2,491.552	2,476.550	1,158.224	1,157.402	1,309.989	1,309.985
σ_ε^2	1.001	0.999	0.987	0.987	1.001	1.002
$l = \ln L$	-8,372.682	-8,356.076	-3,771.467	-3,767.793	-4,610.373	-4,606.511
Ljung-Box	71.263	66.100	78.631	80.143	45.838	43.358
BIC	16,776.640	16,751.247	7,571.251	7,570.978	9,249.446	9,248.892
$\sum_{i \geq 1} \alpha_i + \beta_i$	0.978	1.006	1.015	1.012	1.120	1.245
α_0	14.359	0.494	0.069	0.000	30.885	27.535
$(\hat{\sigma}, t)$	(1.546, 9.3)	(0.114, 4.3)	(0.048, 1.4)	(0.000, 0.0)	(3.600, 8.6)	(2.438, 11.3)
γ		-1.541		1.389		1.374
$(\hat{\sigma}, t)$		(0.204, -7.5)		(0.377, 3.7)		(0.484, 2.8)
α_1	0.492	0.072	0.053	0.049	0.865	0.993
$(\hat{\sigma}, t)$	(0.053, 9.3)	(0.006, 11.9)	(0.006, 8.8)	(0.006, 8.2)	(0.234, 3.7)	(0.133, 7.5)
β_1	0.164	0.329	0.587	0.560	0.000	0.000
$(\hat{\sigma}, t)$	(0.043, 3.8)	(0.062, 5.3)	(0.110, 5.3)	(0.111, 5.1)	(0.063, 0.0)	(0.000, 0.0)
β_2	0.322	0.605	0.375	0.403	0.255	0.252
$(\hat{\sigma}, t)$	(0.033, 9.8)	(0.059, 10.2)	(0.108, 3.5)	(0.108, 3.7)	(0.032, 8.0)	(0.031, 8.1)

TABLE 3.3. MM AGARCH Model Parameters

In order to derive the NIC for MM we replace the parameter values from Table 3.3 into (3.10). Since we will only allow ε_{t-1} to vary we replace all occurrences of $\varepsilon_{t-i}^2, i > 1$ with the expected value $E[\varepsilon_t^2] = 1$ and all h_{t-i} with the sample average $\bar{h}$. Thus for the three different periods we obtain the

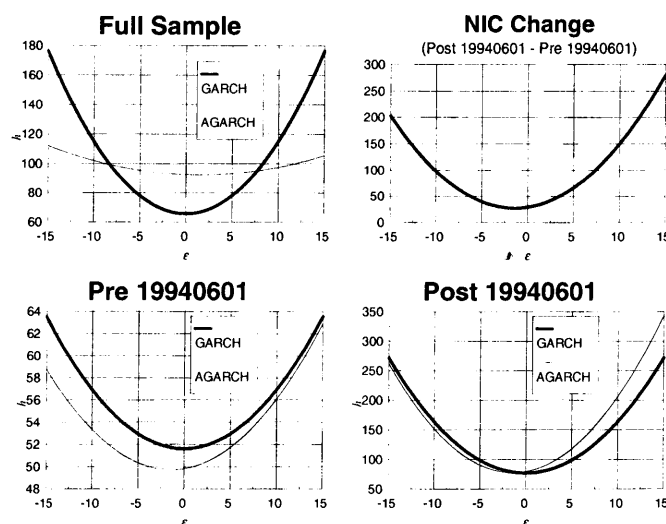


FIGURE 3.10. MM AGARCH News Impact Curve

unconditional NIC

Full Sample GARCH	$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \bar{h} + \beta_2 \bar{h}$ $= 65.808 + 0.492 \varepsilon_{t-1}^2$
Full Sample AGARCH	$h_t = 92.416 + 0.072(\varepsilon_{t-1} - 1.541)^2$
Pre 19940601 GARCH	$h_t = 51.636 + 0.053 \varepsilon_{t-1}^2$
Pre 19940601 AGARCH	$h_t = 49.730 + 0.049(\varepsilon_{t-1} + 1.389)^2$
Post 19940601 GARCH	$h_t = 77.120 + 0.865 \varepsilon_{t-1}^2$
Post 19940601 AGARCH	$h_t = 77.611 + 0.993(\varepsilon_{t-1} + 1.374)^2$

The NIC for these three periods appear in Figure 3.10.

In a way the results for MM was the most disappointing. The AGARCH NIC for the full period is hardly asymmetric and not sensitive to the impact of news. The symmetric GARCH model gives better information as it appears as if the AGARCH model merely fits the mean value for the sample. Although the NIC did increase a lot from the pre 19940601 to the post 19940601 period the asymmetry is still low and even skewed towards positive news for both these periods.

This probably is related to the smooth movements in MM as can be seen in Figure 3.1. The MM is also situated at the short end of the yield curve

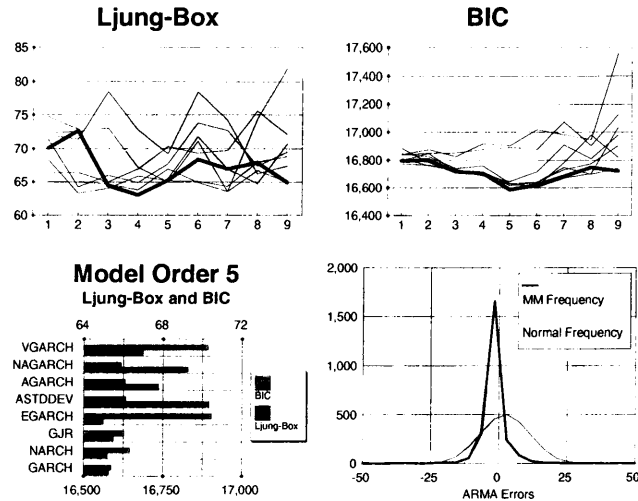


FIGURE 3.11. MM GARCH Models at Order 5

and probably reflects monetary policy more than actual market conditions. As noted earlier the MM did not appear stationary and may require further differencing but this was not performed as it complicates comparisons across markets.

Another possibility is to fit a different model. Previously we indicated that models of order 5 are tentative candidates and if we also evaluate these models we obtain Figure 3.11. Here the GJR(5,5) model based on its relatively low Ljung-Box and *BIC* was chosen

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \gamma S^- \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_5 \varepsilon_{t-5}^2 \\ + \beta_1 h_{t-1} + \cdots + \beta_5 h_{t-5}.$$

Preliminary estimations indicated that α_3, α_4 and α_5 as well as β_1, β_2 and β_3 were close to zero. If we drop the corresponding terms from the model we obtain a modified GJR(2,2) model

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \gamma S^- \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \beta_4 h_{t-4} + \beta_5 h_{t-5}. \quad (3.11)$$

If we fit this model to the different periods we can construct Table 3.4 and from this Figure 3.12. This time the model fits better than before but it is still not without problems. There is a big discrepancy between the

	<i>Full Sample</i>		<i>Pre 19940601</i>		<i>Post 19940601</i>	
	GARCH	GJR	GARCH	GJR	GARCH	GJR
$\bar{y}_t^2$	81.321	81.321	46.088	46.088	120.554	120.554
$\bar{h}_t$	94.863	108.825	51.771	52.029	136.853	132.331
$\sum \varepsilon_t^2$	2,476.678	2,477.941	1,164.729	1,163.274	1,600.116	1,583.424
σ_ε^2	0.998	0.998	0.990	0.989	1.108	1.102
$l = \ln L$	-8,251.240	-8,254.159	-3,725.891	-3,713.942	-4,697.463	-4,692.136
Ljung-Box	65.648	68.709	84.154	87.366	42.987	53.076
<i>BIC</i>	16,541.574	16,555.229	7,487.173	7,470.349	9,430.796	9,427.311
$\sum_{i \geq 1} \alpha_i + \beta_i$	1.010	1.033	1.045	1.032	1.007	0.967
α_0	1.298	1.491	0.197	0.159	1.651	2.619
$(\hat{\sigma}, t)$	(0.177, 7.3)	(0.207, 7.2)	(0.143, 1.4)	(0.121, 1.3)	(0.304, 5.4)	(0.443, 5.9)
γ		0.080		-0.093		0.122
$(\hat{\sigma}, t)$		(0.056, 1.4)		(0.005, -15.5)		(0.086, 1.4)
α_1	0.143	0.127	0.062	0.095	0.154	0.036
$(\hat{\sigma}, t)$	(0.013, 11.4)	(0.017, 7.5)	(0.010, 5.9)	(0.018, 5.3)	(0.018, 8.8)	(0.025, 1.4)
α_2	0.030	0.054	0.103	0.101	0.001	0.022
$(\hat{\sigma}, t)$	(0.007, 4.3)	(0.010, 5.6)	(0.014, 7.4)	(0.014, 7.2)	(0.006, 0.2)	(0.011, 2.0)
β_4	0.406	0.109	0.532	0.554	0.734	0.744
$(\hat{\sigma}, t)$	(0.028, 14.3)	(0.016, 6.9)	(0.042, 12.8)	(0.038, 14.4)	(0.014, 54.0)	(0.015, 50.0)
β_5	0.431	0.703	0.348	0.328	0.118	0.104
$(\hat{\sigma}, t)$	(0.027, 16.1)	(0.015, 46.9)	(0.042, 8.4)	(0.038, 8.6)	(0.013, 9.2)	(0.013, 8.2)

TABLE 3.4. MM GJR Model Parameters

symmetric and asymmetric NICs in the post 19940601 period. This makes it difficult to sensibly interpret the way in which the NIC has changed over time as there is a big difference if the difference in the symmetric NIC is compared to the difference in the asymmetric NIC.

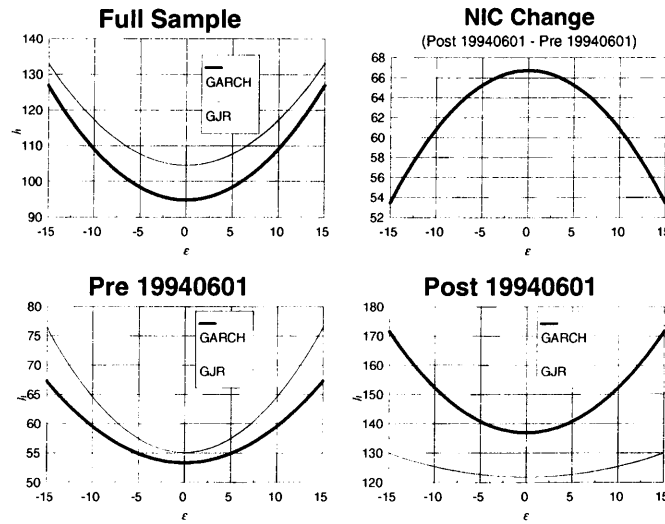


FIGURE 3.12. MM GJR News Impact Curve

The GJR NIC for the different time periods from Table 3.4 is

$$\text{Full Sample GARCH} \quad h_t = 94.832 + 1.298\varepsilon_{t-1}^2$$

$$\text{Full Sample GJR} \quad h_t = 104.576 + 1.491\varepsilon_{t-1}^2 + 0.080S_{t-1}^-\varepsilon_{t-1}^2$$

$$\text{Pre 19940601 GARCH} \quad h_t = 53.360 + 0.197\varepsilon_{t-1}^2$$

$$\text{Pre 19940601 GJR} \quad h_t = 55.082 + 0.159\varepsilon_{t-1}^2 - 0.093S_{t-1}^-\varepsilon_{t-1}^2$$

$$\text{Post 19940601 GARCH} \quad h_t = 136.936 + 1.651\varepsilon_{t-1}^2$$

$$\text{Post 19940601 GJR} \quad h_t = 121.828 + 2.619\varepsilon_{t-1}^2 + 0.122S_{t-1}^-\varepsilon_{t-1}^2.$$

3.3.2. *BM*. Figure 3.13 indicates that either a third or fourth order asymmetric model will suffice for BM. It is also interesting to note the clear under performance of the VGARCH model although some of this may be due to convergence problems as these results are all preliminary.

The GJR(3, 3) model

$$h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \gamma S_{t-1}^-\varepsilon_{t-1}^2 + \alpha_2\varepsilon_{t-2}^2 + \alpha_3\varepsilon_{t-3}^2 \\ + \beta_1 h_{t-1} + \beta_2 h_{t-2} + \beta_3 h_{t-3}$$

was selected for BM based mainly on its relatively low Ljung-Box statistic. In this model both β_1 and β_2 were close to zero and after adjusting the

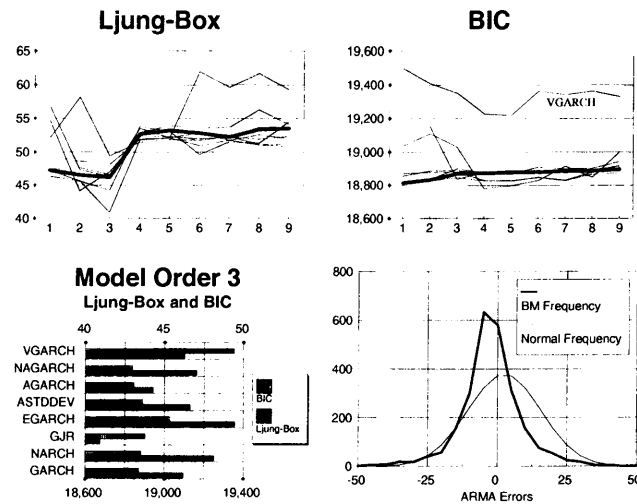


FIGURE 3.13. BM GARCH Models at Order 3

model for this the final BM model is a modified GJR(3,1) model

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \gamma S_{t-1}^- \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \alpha_3 \varepsilon_{t-3}^2 + \beta_3 h_{t-3}. \quad (3.12)$$

We fitted this model for the three different time periods and constructed Table 3.5. The NIC for these periods appear in Figure 3.14. The equations of these curves are

$$\text{Full Sample GARCH} \quad h_t = 138.420 + 0.168 \varepsilon_{t-1}^2$$

$$\text{Full Sample GJR} \quad h_t = 137.960 + 0.130 \varepsilon_{t-1}^2 + 0.092 S_{t-1}^- \varepsilon_{t-1}^2$$

$$\text{Pre 19940601 GARCH} \quad h_t = 75.624 + 0.161 \varepsilon_{t-1}^2$$

$$\text{Pre 19940601 GJR} \quad h_t = 76.279 + 0.187 \varepsilon_{t-1}^2 - 0.083 S_{t-1}^- \varepsilon_{t-1}^2$$

$$\text{Post 19940601 GARCH} \quad h_t = 216.197 + 0.162 \varepsilon_{t-1}^2$$

$$\text{Post 19940601 GJR} \quad h_t = 215.919 + 0.072 \varepsilon_{t-1}^2 + 0.166 S_{t-1}^- \varepsilon_{t-1}^2.$$

If we examine Figure 3.14 we see that the NIC moved upwards quite a bit from the pre 19940601 to the post 19940601 period. The increase is around 140 units of variance or a volatility increase of 12 ($\approx \sqrt{140}$). If we look at Figure 3.1 we see that the pre 19940601 period's movements in BM were less violent than those in the post 19940601 period. The pre

	<i>Full Sample</i>		<i>Pre 19940601</i>		<i>Post 19940601</i>	
	GARCH	GJR	GARCH	GJR	GARCH	GJR
$\bar{y}_t^2$	170.238	170.238	91.375	91.375	242.799	242.799
$\bar{h}_t$	166.848	168.140	90.032	89.410	255.805	255.421
$\sum \varepsilon_i^2$	2,497.626	2,497.661	1,196.105	1,196.095	1,305.795	1,305.795
σ_ε^2	1.001	1.002	1.002	1.003	1.003	1.003
$l = \ln L$	-9,388.249	-9,384.558	-4,291.216	-4,289.489	-5,089.310	-5,083.337
Ljung-Box	46.557	46.831	47.065	47.171	40.234	39.415
BIC	18,815.601	18,816.039	8,617.845	8,621.469	10,214.471	10,209.690
$\sum_{i \geq 1} \alpha_i + \beta_i$	0.875	0.879	0.596	0.619	0.979	0.969
α_0	19.631	19.273	36.232	37.304	10.885	11.238
$(\hat{\sigma}, t)$	(2.265, 8.7)	(2.501, 7.7)	(0.044, 816)	(0.065, 577)	(2.873, 3.8)	(2.840, 4.0)
γ		0.092		-0.083		0.166
$(\hat{\sigma}, t)$		(0.040, 2.3)		(0.045, 1.8)		(0.145, 1.1)
α_1	0.168	0.130	0.161	0.187	0.162	0.072
$(\hat{\sigma}, t)$	(0.021, 7.9)	(0.023, 5.7)	(0.025, 6.6)	(0.031, 6.0)	(0.027, 5.9)	(0.032, 2.2)
α_2	0.119	0.114	0.068	0.068	0.140	0.128
$(\hat{\sigma}, t)$	(0.022, 5.3)	(0.022, 5.2)	(0.025, 2.8)	(0.027, 2.5)	(0.031, 4.5)	(0.029, 4.4)
α_3	0.125	0.117	0.102	0.110	0.143	0.128
$(\hat{\sigma}, t)$	(0.024, 5.2)	(0.024, 4.9)	(0.028, 3.6)	(0.029, 3.8)	(0.031, 4.6)	(0.030, 4.3)
β_3	0.463	0.472	0.265	0.254	0.534	0.558
$(\hat{\sigma}, t)$	(0.040, 11.5)	(0.042, 11.2)	(0.027, 9.7)	(0.036, 7.0)	(0.045, 11.9)	(0.046, 12.2)

TABLE 3.5. BM GJR Model Parameters

19940601 period is also characterised by an extended bull market in South African bonds and this explains why the pre 19940601 NIC actually is more sensitive to good than to bad news. This changes completely in the post 19940601 period where volatility is more sensitive to bad news as is the case for the full sample.

This reflects in part the effect of greater participation of foreigners in the local bond market but is also indicative of the global investment situation.

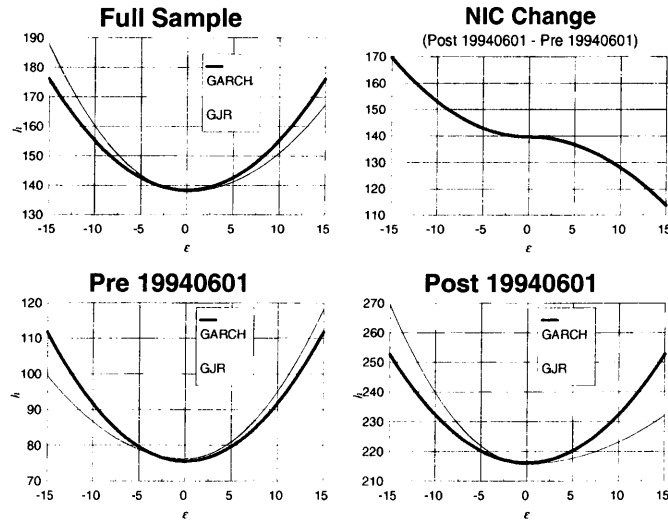


FIGURE 3.14. BM GJR News Impact Curve

During 1994 a number of local incidents added to the volatile bear market such as the resignation of Mr. Keys, then the Minister of Finance, and countless rumours surrounding President Mandela's health. On the global front problems in the Eastern markets resulted in 1998 being a volatile year in emerging and even established markets. This year probably contributes a significant amount to the upward shift in the NIC in the post 19940601 period.

3.3.3. *SM*. The results of the preliminary model estimates for SM is given in Figure 3.15 and it seems appropriate to use a second order and asymmetric model. The ASTDDEV(2,2) model is a candidate based on the combined good performance of the Ljung-Box and *BIC* but this model is symmetric. If we restrict ourselves to asymmetric models the GJR(2,2) and EGARCH(2,2) models appear to be the best. Although the EGARCH(2,2) model has the lowest *BIC* it also has the highest Ljung-Box and the GJR(2,2) model

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \gamma S_{t-1}^- \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \beta_1 h_{t-1} + \beta_2 h_{t-2}$$

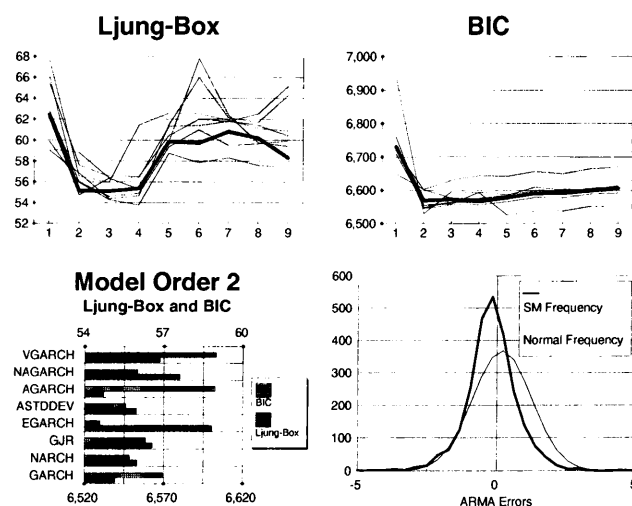


FIGURE 3.15. SM GARCH Models at Order 2

was preferred. Here the α_2 parameter was close to zero and after removing the corresponding term the final model was a GJR(1, 2) model

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \gamma S_{t-1}^- \varepsilon_{t-1}^2 + \beta_1 h_{t-1} + \beta_2 h_{t-2}. \quad (3.13)$$

The model parameters are given in Table 3.6 and the NIC in Figure 3.16. The equations of the NIC are

Full Sample GARCH	$h_t = 1.031 + 0.013\varepsilon_{t-1}^2$
Full Sample GJR	$h_t = 1.039 + 0.014\varepsilon_{t-1}^2 + 0.026S_{t-1}^- \varepsilon_{t-1}^2$
Pre 19940601 GARCH	$h_t = 0.895 + 0.258\varepsilon_{t-1}^2$
Pre 19940601 GJR	$h_t = 0.918 + 0.238\varepsilon_{t-1}^2 + 0.252S_{t-1}^- \varepsilon_{t-1}^2$
Post 19940601 GARCH	$h_t = 1.077 + 0.011\varepsilon_{t-1}^2$
Post 19940601 GJR	$h_t = 1.099 + 0.012\varepsilon_{t-1}^2 + 0.070S_{t-1}^- \varepsilon_{t-1}^2$

If we look at Figure 3.1 we see that the SM is characterised by an extended bull run in the pre 19940601 period. Still the NIC is asymmetric by a much larger degree than in the subsequent post 19940601 period. For the overall period the GJR NIC is hardly asymmetric and quite close to the symmetric GARCH NIC. The NIC did move upwards by a small amount in

	<i>Full Sample</i>		<i>Pre 19940601</i>		<i>Post 19940601</i>	
	GARCH	GJR	GARCH	GJR	GARCH	GJR
$\bar{y}_t^2$	1.094	1.094	1.028	1.028	1.152	1.152
$\bar{h}_t$	1.125	1.131	1.100	1.134	1.216	1.215
$\sum \varepsilon_t^2$	2,509.527	2,508.550	1,194.108	1,194.123	1,315.118	1,314.881
σ_ε^2	1.004	1.004	1.002	1.002	1.005	1.005
$l = \ln L$	-3,257.787	-3,254.932	-1,675.574	-1,664.982	-1,650.881	-1,645.288
Ljung-Box	56.323	56.978	56.791	58.614	57.545	56.402
BIC	6,546.856	6,548.964	3,372.396	3,358.291	3,330.452	3,326.435
$\sum_{i \geq 1} \alpha_i + \beta_i$	0.989	0.879	0.780	1.040	0.997	0.991
α_0	0.013	0.014	0.258	0.238	0.011	0.012
$(\hat{\sigma}, t)$	(0.004, 3.3)	(0.004, 3.3)	(0.052, 5.0)	(0.038, 6.2)	(0.005, 2.2)	(0.004, 2.6)
γ		0.026		0.252		0.070
$(\hat{\sigma}, t)$		(0.004, 6.2)		(0.018, 13.7)		(0.005, 13.8)
α_1	0.084	0.067	0.201	0.076	0.120	0.061
$(\hat{\sigma}, t)$	(0.015, 5.6)	(0.015, 4.5)	(0.038, 5.2)	(0.034, 2.3)	(0.025, 4.9)	(0.023, 2.7)
β_1	0.806	0.836	0.579	0.600	0.620	0.635
$(\hat{\sigma}, t)$	(0.180, 4.5)	(0.182, 4.6)	(0.048, 12.0)	(0.040, 15.0)	(0.128, 4.8)	(0.128, 5.0)
β_2	0.099	0.070			0.257	0.260
$(\hat{\sigma}, t)$	(0.169, 0.6)	(0.171, 0.4)			(0.122, 2.1)	(0.122, 2.1)

TABLE 3.6. SM GJR Model Parameters

the post 19940601 period and it is not difficult to see that this time period was much more volatile than the pre 19940601 period.

3.3.4. *FM*. The results of the preliminary estimations for FM is in Figure 3.17. It is interesting to note the clear distinction between the models in the BIC statistic. Based on this the AGARCH(4,4) model immediately presents itself as the candidate. This model has however the largest Ljung-Box statistic and care is needed. Another alternative is once again the GJR model.

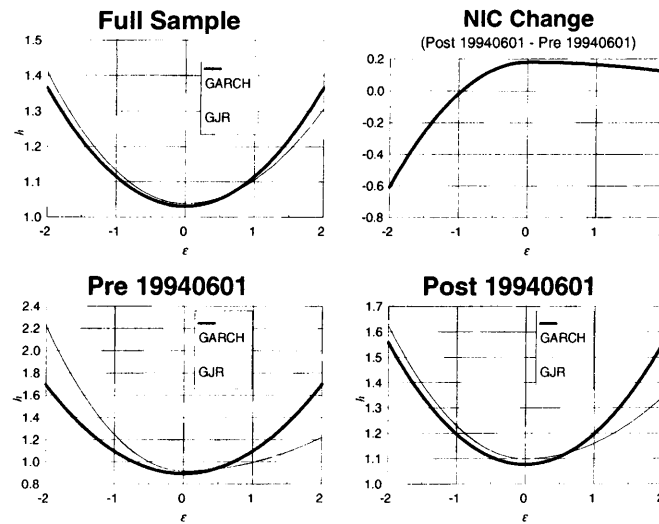


FIGURE 3.16. SM GJR News Impact Curve

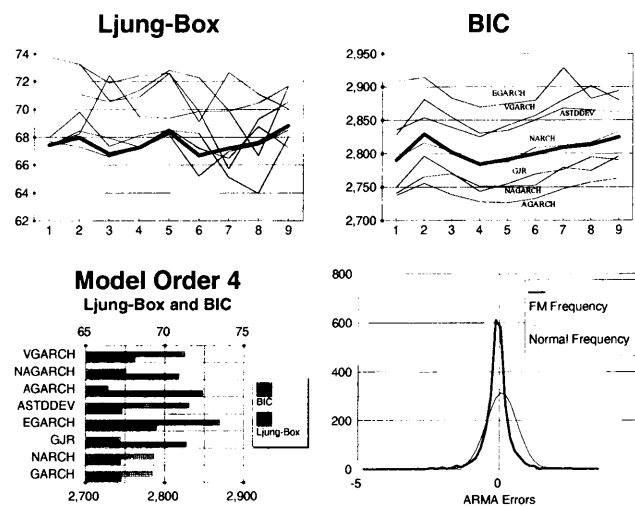


FIGURE 3.17. FM GARCH Models at Order 4

The SAC and SPAC of FM suggested that a fourth order model be used but not necessarily for both the p and q parameters of the model. If the squared errors are modelled using an ARMA model then from the shape of the SAC and SPAC an AR(4) model would also make sense. The coefficients of α_2, α_3 and α_4 were close to zero in the GJR(4,4) model but they were all significant in the AGARCH(4,4) model. If we drop the corresponding terms and model FM using a GJR(1,4) model we should have a better *BIC*

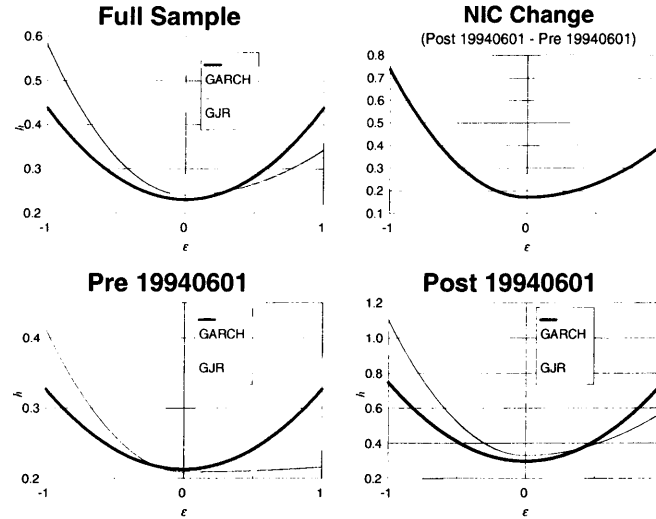


FIGURE 3.18. FM GJR News Impact Curve

due to the much less parameters being used. For this reason the final model selected was such a GJR(1,4) model

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \gamma S_{t-1}^- \varepsilon_{t-1}^2 + \beta_1 h_{t-1} + \beta_2 h_{t-2} + \beta_3 h_{t-3} + \beta_4 h_{t-4}. \quad (3.14)$$

We constructed Table 3.7 after fitting this model to FM and from this the NIC equations are

Full Sample GARCH	$h_t = 0.231 + 0.207\varepsilon_{t-1}^2$
Full Sample GJR	$h_t = 0.241 + 0.102\varepsilon_{t-1}^2 + 0.239S_{t-1}^- \varepsilon_{t-1}^2$
Pre 19940601 GARCH	$h_t = 0.163 + 0.114\varepsilon_{t-1}^2$
Pre 19940601 GJR	$h_t = 0.160 + 0.007\varepsilon_{t-1}^2 + 0.196S_{t-1}^- \varepsilon_{t-1}^2$
Post 19940601 GARCH	$h_t = 0.299 + 0.449\varepsilon_{t-1}^2$
Post 19940601 GJR	$h_t = 0.332 + 0.246\varepsilon_{t-1}^2 + 0.522S_{t-1}^- \varepsilon_{t-1}^2$

This allows us to construct Figure 3.18.

The pre 19940601 NIC is virtually immune to the impact of good news. It is very asymmetric and bad news has a much more significant impact. The asymmetry is characteristic of FM and can be clearly observed in the post 19940601 and the full sample period. The NIC also became even more

	<i>Full Sample</i>		<i>Pre 19940601</i>		<i>Post 19940601</i>	
	GARCH	GJR	GARCH	GJR	GARCH	GJR
$\bar{y}_t^2$	0.292	0.292	0.187	0.187	0.388	0.388
$\bar{h}_t$	0.292	0.312	0.185	0.181	0.473	0.550
$\sum \varepsilon_i^2$	2,499.832	2,500.267	1,196.565	1,199.710	1,302.955	1,302.403
σ_ε^2	1.002	1.002	1.003	1.005	1.002	1.002
$l = \ln L$	-1,356.833	-1,328.216	-595.288	-573.431	-735.924	-712.630
Ljung-Box	67.649	72.150	65.306	76.199	57.611	59.082
<i>BIC</i>	2,760.589	2,711.173	1,233.066	1,196.428	1,514.865	1,475.440
$\sum_{i \geq 1} \alpha_i + \beta_i$	0.957	0.955	0.927	0.893	1.056	1.092
α_0	0.012	0.012	0.013	0.017	0.012	0.010
$(\hat{\sigma}, t)$	(0.002, 7.2)	(0.002, 7.7)	(0.004, 3.4)	(0.004, 4.1)	(0.002, 6.6)	(0.002, 6.1)
γ		0.239		0.196		0.522
$(\hat{\sigma}, t)$		(0.005, 47.1)		(0.009, 20.8)		(0.002, 259)
α_1	0.207	0.102	0.114	0.007	0.449	0.246
$(\hat{\sigma}, t)$	(0.026, 7.7)	(0.020, 5.1)	(0.021, 5.5)	(0.021, 0.3)	(0.077, 5.8)	(0.048, 5.1)
β_1	0.523	0.456		0.532	0.326	0.254
$(\hat{\sigma}, t)$	(0.107, 4.9)	(0.096, 4.7)		(0.182, 2.9)	(0.106, 3.1)	(0.080, 3.2)
β_2	0.000	0.000	0.344	0.000	0.000	0.009
$(\hat{\sigma}, t)$	(0.128, 0.0)	(0.106, 0.0)	(0.127, 2.7)	(0.279, 0.0)	(0.101, 0.0)	(0.070, 0.1)
β_3	0.000	0.042		0.001	0.238	0.233
$(\hat{\sigma}, t)$	(0.075, 0.0)	(0.089, 0.5)		(0.230, 0.0)	(0.121, 2.0)	(0.110, 2.1)
β_4	0.227	0.235	0.469	0.255	0.043	0.089
$(\hat{\sigma}, t)$	(0.075, 3.0)	(0.067, 3.5)	(0.116, 4.0)	(0.159, 1.6)	(0.088, 0.5)	(0.083, 1.1)

TABLE 3.7. FM GJR Model Parameters

sensitive to bad (and good) news in the post 19940601 period. Figure 3.1 clearly displays the problems that the Rand has faced in recent times and, if these are compared to the relative stability of the earlier pre 19940601 period, it also explains why the NIC has changed so much.

4. Conclusion

Returns in the various markets are of a volatile nature (see Table 3.1). The transformation applied in MM and BM as well as in SM and FM is the same and we can compare these markets to one another. The risk in BM is much higher than that in MM which is to be expected since investments in BM are of a much longer duration than those in MM. The higher risk observed in SM may be due to the much lower daily turnover in this market compared to FM.

Returns in all markets are negative skew which is also indicative of an asymmetric NIC. The kurtosis of all markets are also much higher than that of a normal distribution and fat tailed leptokurtic models will be required. Daily returns are also fairly random and even though significant autocorrelations and partial autocorrelations were observed ARMA models of a fairly high order still do not have any real prediction power.

The R^2 statistic for all markets were very low – not more than 10% for any market. The highest R^2 was observed in the MM which could be indicative of the market's ability to preempt monetary policy as well as the trending nature thereof. The lowest R^2 was observed in the BM and then the FM. These markets have the highest volumes and it could be indicative of near perfect market conditions (in terms of the efficient market hypothesis).

In modelling market volatility the GJR model was the asymmetric GARCH model of choice for almost all the markets evaluated here. This corresponds to the findings of a similar study conducted by Engle and Ng (1993) where the NIC for Japanese stock returns were constructed using the following models:

- GARCH(1, 1)
- EGARCH(1, 1)
- AGARCH(1, 1)
- VGARCH(1, 1)
- NAGARCH(1, 1)
- GJR(1, 1).

Of the four markets evaluated the MM presented most problems and it should also be treated differently to the rest as movements in this market often reflects monetary policy rather than supply and demand. South Africa has recently adopted a policy of market determined short term rates and it will be interesting to observe the NIC for the MM market in the future.

The other three markets, BM, SM and FM, mostly had negative asymmetric NICs. The BM had a positive asymmetric NIC in the pre 19940601 period and negative asymmetric thereafter. It is interesting to note that in the pre 19940601 period the BM experienced an extended bull market but that the post 19940601 period is more representative and reflects both bull and bear markets. This may explain why the pre 19940601 NIC is positive asymmetric.

Both the SM and FM had negative asymmetric NICs in the pre and post 19940601 periods with the FM becoming more and the SM less asymmetric. Here again the SM is characterised by an extended bull run in the pre 19940601 period while the post 19940601 sample is much more volatile.

Volatility in general were higher in all four markets in the latter part of the sample and this could be the result of a less controlled market that is heavily dependent on foreign capital and sensitive to movements in these capital flows. It could also reflect a more volatile period in global markets experienced in recent times. Whichever explanation is the true one, volatility is on the rise and this demands a theoretical framework able to model this important phenomena!

Bibliography

- [1] L.J. Bain and M. Engelhardt: *Introduction to probability and mathematical statistics*, PWS (1987) ISBN 0-87150-067-1
- [2] F. Black: *Studies in stock price volatility changes*, Proceedings of the 1976 Business Meeting of the Business and Economics Statistics Section, American Statistical Association (1976) pp. 177-181.
- [3] T. Bollerslev: *Generalised autoregressive conditional heteroskedasticity*, Journal of Econometrics **31** (1986) pp. 307-327.
- [4] T. Bollerslev, R. Chou and K. Kroner: *ARCH modelling in finance: A review of the theory and empirical evidence*, Journal of Econometrics **52** (1992) pp. 5-59.
- [5] B.L. Bowerman and R.T. O'Connell: *Time series forecasting unified concepts and computer implementations*, Second edition, Duxbury (1987) ISBN 0-87150-070-1
- [6] G.E.P. Box and D.R. Cox: *An analysis of transformations (with discussion)*, Journal of the Royal Statistical Society B. **26** (1964) pp. 211-252.
- [7] G.E.P. Box and G.M. Jenkins: *Time series analysis forecasting and control*, Revised edition, Holden-Day (1976) ISBN 0-8162-1104-3.
- [8] J.Y. Campbell, A.W. Lo and A.C. MacKinlay, *The econometrics of financial markets*, Princeton (1997) ISBN 0-691-04301-9.
- [9] N.E. Coulson and R.P. Robins, *Aggregate economic activity and the variance of inflation: another look*, Economics Letters **17** (1985) pp. 71-75.
- [10] B. de la Rosa et al.: *Inleidende Algebra*, McGraw-Hill (1987) ISBN 0-07-450695-1.
- [11] P.J. Dhrymes: *Econometrics: statistical foundations and applications*, Study edition, Springer-Verlag (1984) ISBN 0-387-90095-0.
- [12] N.R. Draper and H. Smith: *Applied regression analysis*, John Wiley & Sons (1980) ISBN 0-471-02995-5.
- [13] J. Durbin: *Estimation of parameters in time-series regression models*, Journal of the Royal Statistical Society B. **22** (1960) pp. 139-153.
- [14] R. Ellis and D. Gulick: *Calculus with analytical geometry*, Third edition, Harcourt Brace Jovanovich (1986) ISBN 0-15-505737-5.
- [15] R.F. Engle: *Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom Inflation*, Econometrica **50** No. 4 (July 1982) pp. 987-1007.
- [16] R.F. Engle: *Statistical models for financial volatility*, Financial Analysts Journal (January-February 1993) pp. 72-78.

- [17] R.F. Engle and V.K. Ng: *Measuring and testing the impact of news on volatility*, The Journal of Finance **XLVII** No. 5 (December 1993) pp. 1749-1778.
- [18] J.B. Fraleigh and R.A. Beauregard: *Linear Algebra*, Addison-Wesley (1987) ISBN 0-201-15449-8.
- [19] W.A. Fuller: *Introduction to statistical time series*, John Wiley & Sons (1976).
- [20] L. Glosten, R. Jaganathan and D. Runkle: *Relationships between the expected value and the volatility of the nominal excess return on stocks*, Columbia University, Department of Finance Working Paper (1989).
- [21] C.W.J. Granger and P. Newbold: *Forecasting economic time series*, Second edition, Academic Press (1986) ISBN 0-12-295184-0.
- [22] D.N. Gujarati: *Basic Econometrics*, McGraw-Hill (1988) ISBN 0-07-100446-7.
- [23] J.D. Hamilton: *Time series analysis*, Princeton (1994) ISBN 0-691-04289-6.
- [24] M. Kendall and J.K. Ord: *Time series*, Third edition, Edward Arnold (1990) ISBN 0-340-59327-X.
- [25] D.E. Knuth: *The art of computer programming*, Second edition, Fundamental Algorithms 1 Addison-Wesley (1973).
- [26] E. Kreyszig: *Introductory functional analysis with applications*, John Wiley & Sons (1978).
- [27] J.H. Matthews: *Numerical methods for computer science, engineering, and mathematics*, Prentice-Hall (1987) ISBN 0-13-626565-0.
- [28] D. Meyers: *The discrete Fourier transform illusion*, Technical analysis of stocks and commodities (April 1999) pp.16-27.
- [29] D. Meyers: *The endpoint fast Fourier transfer system*, Technical analysis of stocks and commodities (May 1999) pp.26-40.
- [30] D.F. Morris: *Multivariate statistical methods*, International edition, McGraw-Hill (1990) ISBN 0-07-043187-6.
- [31] C.R. Nelson and H. Kang: *Spurious periodicity in inappropriately detrended time series*, Econometrica **49** (1980) pp. 741-752.
- [32] M.J.D. Powell: *Approximation theory and methods*, Cambridge (1981) ISBN 0-521-29514-9.
- [33] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling: *Numerical recipes in C*, Cambridge (1989) ISBN 0-521-35465-X.
- [34] J.N. Rayner: *An introduction to spectral analysis*, Pion (1971).
- [35] A. Schuster: *On the investigation of hidden periodicities with application to a supposed 26-day period of meteorological phenomena*, Terrestrial Magnetism **3** (1898) pp. 13-41.
- [36] E. Slutsky (1927): *The summation of random causes as the sources of cyclic processes*, (Russian) English translation, Econometrica **5** (1937) pp. 105-146.

- [37] T. Teräsvirta: *Two stylized facts and the GARCH(1,1) model*, Stockholm School of Economics, Working Paper **96** (1996).
- [38] T. Teräsvirta: *Properties of moments of a family of GARCH processes*, Stockholm School of Economics, Working Paper **198** (1997).
- [39] M. Townsend: *Discrete mathematics: applied combinatorics and graph theory*, Benjamin/Cummings (1987) ISBN 0-8053-9355-2.
- [40] A.A. Weiss, *ARMA models with ARCH errors*, Journal of Time Series Analysis **5** (1984) pp. 129-143.
- [41] R.J. Wonnacott and T.H. Wonnacott, *Econometrics*, John Wiley & Sons (1970).
- [42] G.U. Yule: *On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers*, Phil. Trans. A. **226** (1927) pp. 267-298.

/

Subject and Notation Index

<i>Symbol</i>	<i>Description</i>	<i>Page</i>
	absolute summable	7
AIC	Akaike's information criterion	135
d	amplitude	24
AGARCH	asymmetric GARCH	124
ρ_k	autocorrelation	15
ACF	autocorrelation function	19
AR	autoregressive	55
ARCH	autoregressive conditional heteroscedastic	107
ARIMA	autoregressive integrated moving average	100
ARMA	autoregressive moving average	81
B	backward shift operator	9
$[0, 2\pi]$	basic interval	23
BIC	Bayesian information criterion	135
BM	bond market	129
$y_t^{(\lambda)}$	Box-Cox transformations	104
h_t	conditional variance	107
CPI	consumer price index	12
	covariance stationary	3
$F(y, \theta)$	cumulative probability density function	90
$\frac{\partial y_t}{\partial \varepsilon_{t-k}}$	dynamic multiplier	36
	difference stationary	10
FM	forex market	129
f	frequency	24
GARCH	generalized autoregressive conditional heteroscedastic	114
GJR	Glosten, Jagannathan and Runkle	124

<i>Symbol</i>	<i>Description</i>	<i>Page</i>
	headline inflation	12
	homoscedastic	103
	heteroscedastic	103
i.i.d.	identically and independently distributed	6
	inflation	12
I	information matrix	133
Ψ_t	information set	106
	integrated models	98
$I_n(\lambda)$	intensity	28
L	likelihood function	131
Q^*	Ljung-Box statistic	134
l	log likelihood function	133
$\frac{\partial \sum_{i=0}^{\infty} y_{t+i}}{\partial \varepsilon_t}$	long run effect	48
μ_r	moment about the mean	88
μ'_r	moment about the origin	88
$M_y(k)$	moment generating function	88
	moments	87
MM	money market	129
MA	moving average	78
MARCH	multiplicative ARCH	124
NARCH	nonlinear ARCH	124
NAGARCH	nonlinear asymmetric GARCH	124
$N(\mu, \sigma^2)$	normal distribution	90
NIC	news impact curve	126
$\frac{\pi}{\Delta t}$	Nyquist frequency	29

<i>Symbol</i>	<i>Description</i>	<i>Page</i>
	random walk	58
ϕ_{kk}	partial autocorrelation	19
PAC	partial autocorrelation function	21
p	period	24
	phase angle	24
θ	phase shift	24
$y_t^{(\lambda)}$	power transformation	104
$f(y, \theta)$	probability density function	90
SAC	sample autocorrelation function	19
SPAC	sample partial autocorrelation function	21
	Schuster periodogram	28
BIC	Schwartz' Bayesian information criterion	135
	second order stationary	3
SM	share market	129
	spectrogram	28
	square summable	8
	stable	47
$N(0, 1)$	standard normal distribution	90
	stationary	3
	strictly stationary	3
S	sum of squared errors	131
$\mathcal{S}$	$\frac{\partial^2 S}{\partial \theta \partial \theta'}$	134
y_t	time series	1
	trend stationary	10
y_t	unexpected return	125
VGARCH	VGARCH	124
	weakly stationary	3
ε_t	white noise	6
z_t	zero mean time series	52

Abstract

Supervisor Prof F.D. van Niekerk
Department Mathematics and Applied Mathematics
Degree MSc

We investigate the issues surrounding linear time series models. Our specific interest is in stationary time series models and difference equations. We derive an analytical expression for the inverse of the matrix $\mathbf{F}$ characterising such difference equations using Vandermonde's matrix theory. We use these results to discuss the homoscedastic ARMA and heteroscedastic ARCH and GARCH type time series models. Finally we use the ARCH and GARCH type models to construct various asymmetric news impact curves for South African financial time series.

Samevatting

Leier Prof F.D. van Niekerk
Departement Wiskunde en Toegepaste Wiskunde
Graad MSc

Ons ondersoek lineêre tydreeks modelle en stel veral belang in stasionêre tydreeks modelle sowel as verskil vergelykings. Ons lei analitiese uitdrukkings af vir die inverse van die $\mathbf{F}$ matriks wat sulke verskil vergelykings beskryf deur gebruik te maak van Vandermonde se matriks teorie. Ons gebruik hierdie resultate om die homoskedastiese ARMA en die heteroskedastiese ARCH en GARCH tipe modelle te ondersoek. Laastens gebruik ons die ARCH en GARCH tipe modelle om verskeie ongelykmatige nuus sensitiviteits kurwes vir Suid Afrikaanse finansiële tydreeks te konstrueer.