

Panel data techniques and accounting research

P de Jager

**Department of Financial Management
University of Pretoria**

Abstract

Empirical accounting research frequently makes use of data sets with a time-series and a cross-sectional dimension – a panel of data. The literature review indicates that South African researchers infrequently allow for heterogeneity between firms when using panel data and the empirical example shows that regression results that allow for firm heterogeneity are materially different from regression results that assume homogeneity among firms. The econometric analysis of panel data has advanced significantly in recent years and accounting researchers should benefit from those improvements.

Key words

Data panel
Fixed effects
Heterogeneity
Panel data
Pooling
Poolability
Random effects

Acknowledgement

The assistance of Prof. R. van Eyden of the Economics Department at the University of Pretoria is gratefully acknowledged.

1 Introduction

South African empirical accounting studies often make use of accounting data obtained from locally listed companies. These data are usually annual or at best semi-annual – hence the difficulty of collecting a statistically large enough sample.

According to Lind, Marchal and Wathen (2005:264), statisticians consider a sample of 30 or more observations to be large enough for the central limit theorem to be employed when the distribution of a population is unknown (for South African listed companies this implies 30 years of observations for annual data). Practical difficulties in collecting and using a time series of data of this length include the structural break in the South African economy around 1994 (end of the Apartheid era), the fact that few of the companies


currently listed on the Johannesburg Stock Exchange (JSE) have been in existence for 30 years¹ and the large number of financial accounting changes that took place in the 1990s with the harmonisation/internationalisation of South African accounting practices (Steyn & Hamman 2004:130).

Researchers frequently compensate for the lack of time depth in their data series by also collecting data cross-sectional for different companies. Examples include De Wet and Du Toit (2007) who used a dataset with a time depth of 11 years across 83 companies to investigate the relationship between shareholder returns and different internal performance measures, and De Jager and De Wet (2007) who used a dataset with a time depth of 13 years across 53 companies to investigate the relationship between market value added (MVA) and five selected performance measures.

Figure 1 Time-series data, cross-sectional data and panel data


Example of: time-series data

Company 1	
Year	ROE
1995	12%
1996	10%
1997	23%
1998	9%
1999	12%



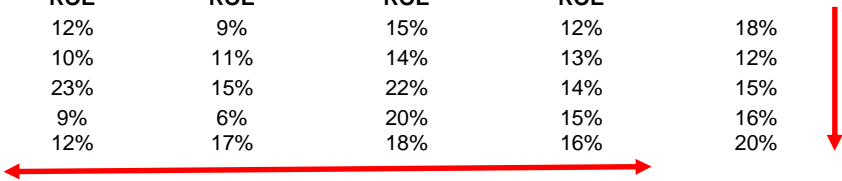
Example of cross-sectional data:

	Company 1	Company 2	Company 3	Company 4	Company 5
Year	ROE	ROE	ROE	ROE	ROE
1995	12%	9%	15%	12%	18%



Example of panel data:

	Company 1	Company 2	Company 3	Company 4	Company 5
Year	ROE	ROE	ROE	ROE	ROE
1995	12%	9%	15%	12%	18%
1996	10%	11%	14%	13%	12%
1997	23%	15%	22%	14%	15%
1998	9%	6%	20%	15%	16%
1999	12%	17%	18%	16%	20%



In both of the above examples of accounting researchers using panel data, the data were stacked and a single intercept coefficient and a single slope coefficient for each of the explanatory variables estimated. The above method is consistent with the complete pooling of the panel, based on the assumption that firms in the sample are homogeneous. This method neglects to make use of the advances available in the statistical analysis of panel data.

¹ Only 24% of companies currently listed (April 2008) have accounting time series dating back 30 years.

Figure 2 Stacked data and a simple regression thereof

		Y (market return)	X (ROE)
Company 1	1995	10%	12%
	1996	12%	10%
	1997	15%	23%
	1998	15%	9%
	1999	20%	12%
Company 2	1995	14%	9%
	1996	20%	11%
	1997	19%	15%
	1998	14%	6%
	1999	15%	17%
Company 3	1995	12%	15%
	1996	16%	14%
	1997	17%	22%
	1998	15%	20%
	1999	18%	18%

Estimation result, using all data, pooled: $Y = aX + c = 0.093X + 0.141$

2 Research objectives

The objectives of this paper are twofold. The first objective is to show that South African accounting researchers seldom use the econometric advances available in the analysis of panel data. This will be done by contrasting South African accounting research that makes use of panel data with international accounting research that employs panel data.

The second objective of this paper is to illustrate the misleading inference that may result from the use of inappropriate estimation methodologies in panel settings. This will be done by providing a practical illustration of the econometric issues discussed in the study. A regression model will be built in which the market value added (MVA) of 53 companies for 13 years is explained by accounting variables. The vantage point will be a pooled regression in which a common intercept and slope coefficient will be calculated for each of the explanatory variables. Guided by relevant statistical tests, the regression model will be adjusted to a panel data appropriate model. The final regression results will be compared with the initial regression result and the material differences highlighted.

3 Literature review

The first section of the literature review will be used to establish, in a generic sense, what econometric technology is available for analysing panel data. The second section of the literature review will deal specifically with the analysis of panel data in an accounting research environment, focusing on current practices in South Africa as opposed to international practices.

3.1 Panel data and its analysis

The term "panel data" refers to the pooling of observations on a cross-section of, say, firms, countries, etc., over several time periods (Baltagi 2005:1). Creating a data panel or

“pooling” data (as it is also known) can be used to compensate for a lack of time-series depth available in data where it can increase degrees of freedom and potentially lower standard errors of the coefficients of a regression.

The earliest panel specifications emphasised the joint estimation of coefficients such as the example from the airline industry in Greene (2008:286):

$$\ln cost_{it} = \beta_1 + \beta_2 \ln output_{it} + \beta_3 \ln fuel\ price_{it} + \beta_4 \ln load\ factor_{it} + \varepsilon_{it}$$

The panel was constructed using data from 25 firms over 15 years and ordinary least squares (OLS) estimation was employed to obtain a single intercept coefficient and a single slope coefficient for each of the explanatory variables. In other words, the data from each firm were simply stacked together and the pooled regression coefficients estimated (see the example in figure 2). Restricting the specification to a single intercept coefficient and a single slope coefficient for each of the explanatory variables fails to account for any heterogeneity between airlines in this example (or firms in the example in figure 2).

Greene (2008:334) concludes that the main advantage of panel data is that one can formally model the heterogeneity across groups that are typically present in panel data. Baltagi (2005:4) confirms this in his statement that the first benefit of panel data is “controlling for individual heterogeneity”. Additional benefits of using panel data include the following:

- Panel data provide more informative data, more variability, less collinearity among the variables, more degrees of freedom and more efficiency.
- Panel data are better able to study the dynamics of adjustment.
- Panel data are better able to identify and measure effects that are simply not detectable in pure cross-section or pure time-series data.
- Panel data allow for the construction and testing of more complicated behavioural models than purely cross-section or time-series data. Using panel data may eliminate biases resulting from aggregation over firms or individuals.

3.1.1 *The one-way error component regression model*

The one-way error component regression model allows for heterogeneity in the error term in terms of (a) the specific cross-section (or the specific firm in the example in figure 2); or (b) the specific time period.

In vector notation:

$$\gamma_{it} = \alpha + X_{it}'\beta + u_{it} \quad (i = 1, \dots, N; t = 1, 2, \dots, T)$$

with

$$u_{it} = \mu_i + v_{it} \quad (\text{this implies a unique intercept coefficient for each cross-section or firm in the above example})$$

$$\text{or } u_{it} = \lambda_t + v_{it} \quad (\text{this implies a unique intercept coefficient for each time period})$$

v_{it} is independent and identically distributed $\text{IID}(0, \sigma_v^2)$ – a “well-behaved” remainder disturbance term. N represents the number of cross-sections and T the number of time periods.

Different assumptions about μ_i (or λ_t) lead to either using the fixed effects approach or the random effects approach when estimating the different coefficients².

In terms of the fixed effects approach, μ_i are assumed to be fixed parameters to be estimated, and the observations of the exogenous variables X_{it} are assumed to be independent of the error term v_{it} for all cross-sections or time periods. According to Baltagi (2005:12), this is an appropriate specification if one is focusing on a specific set of firms and inference is limited to that set of firms – that is, this is an appropriate specification form for most accounting research.

According to the random effects approach, the μ_i parameters are assumed to be random IID($0, \sigma_\mu^2$), v_{it} is IID($0, \sigma_v^2$) and μ_i are independent of v_{it} . In addition, the total error and the observations of the exogenous variables are assumed to be independent for all cross-sections or time periods. The random effects approach is appropriate when drawing “random” draws from a large population (N approaching infinity) to make inferences about the characteristics of the population. This renders the random effects approach inappropriate for the majority of accounting research projects. Hence the remainder of the literature review on panel data and its analysis will focus on the fixed effects approach.

The fixed effects approach

The fixed effects approach assumes that differences across cross-sections can be captured in differences in the constant term. In the model outlined above, this difference in the constant term is included in the error term as either μ_i (for a fixed cross-sectional effect) or λ_t (for a fixed time effect). The parameters of the model can be estimated using ordinary least squares including dummy variables for each cross-section (or time period), the LSDV method; or the data can be demeaned to “wipe” out any individual effects. β is estimated and the individual effects are then calculated per cross-section (or time period) using first-order conditions – the so-called “Within” method (Baltagi 2005:13).

The LSDV model is appealing, but a large number of parameters must be estimated (coefficients for all the X-regressors plus the intercept plus a coefficient for each cross-section [or time period]) and degrees of freedom problems may result, especially if N is large. In such instances, the “Within” method can be used, but note that because the cross-sectional constant terms are calculated and not estimated in the model, no statistical inference can be drawn from individual (or time) effects since no standard errors or t-statistics are available.

According to Baltagi (2005:13), the fixed effects approach suffers from consistency problems. If the number of time periods used in the data panel is fixed (small) and the number of cross-sections approaches infinity, then only the fixed effects estimation of β is consistent, while the fixed effects estimation of the individual cross-section intercepts is not.

However, for the accounting researcher it is more important to know that not controlling for individual fixed effects in a data panel can lead to an omitted variable bias³ problem and inconsistent estimates of the regression parameters. It is to be expected in an accounting data panel that there will be heterogeneity between different companies or cross-sections

² μ_i or λ_t can be regarded as interchangeable concepts in the discussion on the choice between the fixed effects approach or the random effects approach.

³ Some texts use the term “omission variable bias”.

and therefore that estimating a regression equation that does not account for individual differences will lead to biased and inconsistent results.

The above expectation can also be statistically tested and confirmed by means of a simple Chow test with the H_0 that the intercept is the same across all cross-sections. The F-test statistic is calculated and the critical value of the test statistic is $F_{(N-1),(NT-N-K)}$. If the calculated F-value exceeds the critical F-value, the null hypothesis of a common intercept for all cross-sections is rejected and the data should not be pooled for regression purposes. Not controlling for this heterogeneity when the test indicates heterogeneity will lead to biased and inconsistent parameter estimates.

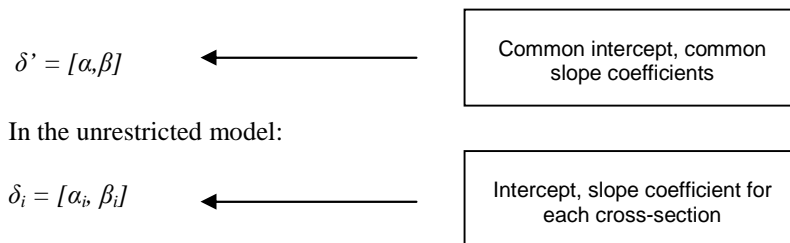
A simple F-test can also be used to test for the validity of fixed time effects, that is, heterogeneity between time periods. A two-way error component regression model allows for heterogeneity in the error term in terms of firstly, the specific cross-section (or the specific firm in the example in figure 2), and secondly, the specific time period. An applied F-test can be used to test the H_0 of a common intercept across time and another common intercept across cross-sections versus the H_A of a unique intercept for each year and cross-section.

The next section deals with more advanced hypothesis testing with fixed effects in the one- and two-way error component models.

3.1.2 Hypothesis testing with panel data

Testing for poolability

Poolability refers to the calculation of a common slope and a common intercept across all cross-sections. The more restrictive definition of poolability is that all coefficients (slopes and intercepts) are the same across time and cross-sections. In the unrestricted model, slope and intercept coefficients are allowed to vary across time and cross-sections. In the restricted model:



The H_0 for the general poolability test is that $\delta_i = \delta$ for all i . The H_A is that the δ_i are not all equal to δ .

An F-test is used to test the above. The critical value for the F-statistic is defined as $F_{((N-1)K', N(T-K'))}$. If the calculated F-value exceeds the critical F-value, then the null hypothesis of a common intercept and common slope coefficients between the cross-sections is rejected and the data should not be pooled for regression purposes. Here no conclusion can be drawn about the validity of fixed or random effects.

Serial correlation

Serial correlation refers to the situation in which residuals are correlated across time. Ignoring serial correlation where it exists, causes consistent but inefficient estimates and

biased standard errors. Inference about the significance of X-regressors may be incorrect under serial correlation conditions.

Testing for serial correlation, given fixed effects, is done by means of $H_0: \rho = 0$ and $H_A: |\rho| > 0$. ρ is a linear approximation of the relationship between the current and the previous period error terms.

An LM test statistic is calculated and asymptotically distributed $N(0,1)$. The critical value for the LM statistic is $N(0,1)$. If the calculated LM value exceeds the critical LM value then the null hypothesis of no serial correlation is rejected and serial correlation is present. Serial correlation can be corrected for by using the Prais-Winston transformation.

Heteroscedasticity

In the previous models we assumed that disturbances are homoscedastic with the same variances across time and cross-sections. According to Baltagi (2005:79), this may be a restrictive assumption for data panels, where cross-sectional units may often be a different size and as a result exhibit different variations. Assuming homoscedastic disturbances when heteroscedasticity is present will yield consistent estimation results of coefficients that are not efficient. The standard errors of the estimates will be biased and the inference about the significance of X-regressors may be incorrect.

Testing for heteroscedasticity is done with $H_0: \sigma_i^2 = \sigma^2$ for all i and H_A : not equal for all i where σ_i^2 represents the variance of the error terms of cross-section i .

An LM statistic is calculated to test the above and the critical value for the LM statistic is $\chi^2_{(N-1)}$. If the calculated LM value exceeds the critical value, the null hypothesis of homoscedasticity is rejected and heteroscedasticity in the error terms seems highly likely and will have to be corrected.

Misspecification of regressors

The random effects model assumes exogeneity of all the regressors with the random individual effects (Baltagi 2005:19). If this exogeneity assumption of the random effect model is not valid, then the GLS regression results will become biased and inconsistent.

In the previous part of the discussion it was pointed out that the majority of accounting researchers tend to use the fixed effects approach where endogeneity of the regressors is not such a serious problem. This section will thus conclude by stating that where the Hausman test for endogeneity rejects the H_0 of no endogeneity of regressors, this can be compensated for by using the Seemingly Unrelated Regressions (SUR) technique to estimate the parameters.

SUR

The SUR technique can thus be used to overcome the problem of endogenous regressors. This technique models endogenous variables as linear functions of lagged endogenous variables and all exogenous variables in the system. The system of equations in the model is estimated jointly. This means that the effect of the independent variables on each endogenous variable takes into account the endogenous nature of the other endogenous variables.

SUR techniques are also useful because they allow for slope coefficients and the intercept coefficients to vary across cross-sections. In the previous approaches discussed,

allowance was only made for intercept coefficients to vary across cross-sections. This captures efficiency due to the correlation of disturbances across equations.

3.2 Accounting research and panel data

The objective of the following section of the literature review is to demonstrate the lack of the use of panel data techniques by South African accounting researchers when regressing data with a time and a cross-sectional dimension, and not to provide a complete literature review of the use of panel data techniques in accounting research.

The choice of local and international journals to include in the review was limited to accounting journals to isolate the use of panel data techniques in an accounting research context. South African journals included *Meditari Accountancy Research*, the *South African Journal of Accountancy Research* and the *South African Journal of Business Management*. International journals included *The Accounting Review*, the *Journal of Accounting and Economics* and the *Journal of Accounting Research*.

Electronic copies of these journals were searched for the use of the term “regression”, and the data sets used in the identified studies evaluated for a panel data structure (time-series and cross-sectional properties). The techniques used in the articles identified were evaluated and are discussed below. Preference was given to more recent articles and the author used his judgement in selecting which articles to include.

3.2.1 South Africa journals

Meditari Accountancy Research

One instance was found where researchers made use of panel data regression techniques. Swartz, Swartz and Firer (2006) used a panel data set consisting of accounting and other variables from 154 firms over a period of eight years. Their model specification and results indicate that they employed a one-way error component regression model using a fixed effects approach. They regressed a market variable (price) on mostly accounting variables and the resultant R^2 of 0.9114 (Swartz *et al.* 2006:76) would not have been possible without the unique intercept term per firm. A shortcoming of the empirical work performed was that no hypothesis tests were conducted to support the statistical validity of results. These include tests for poolability of data, tests for the validity of fixed effects, tests for serial correlation and tests for heteroscedasticity.

A study by De Wet (2005) utilised a panel of 89 companies [over a period of 11 years. The data were simply pooled and a common slope and intercept coefficients estimated. Heterogeneity between firms was not controlled for and the study may have suffered from omitted variable bias. Firer and Stainbank (2003), Gouws and Lucouw (2000) and Swartz and Firer (2005) made use of a cross-section of data in the empirical part of their research (one year of data by multiple firms). The use of panel data would have had benefits such as more reliable parameter estimates, more degrees of freedom, and importantly, the ability to study the dynamics of adjustment from one period to another.

South African Journal of Accounting Research

No instance was found in which researchers utilised panel data regression techniques. Negash (2001) used a panel data set consisting of market data, accounting data and other variables from 51 firms over a period of 52 weeks. All indications are that the data were

simply pooled to obtain a common slope and intercept coefficients. Heterogeneity between firms was not controlled for and the study may have suffered from omitted variable bias.

Van Staden (1999) used a panel data set consisting of market and accounting variables from 150 firms over a period of five years. Numerous regressions were estimated including a simple pooled model in which common slope and intercept coefficients were estimated. Ad hoc provision was made for the unique panel characteristics of the data by checking the stability of the estimates per period and per sector. Heterogeneity between firms was not controlled for and the study may have suffered from omitted variable bias.

South African Journal of Business Management

De Wet and Du Toit (2007) used a panel data set with a time depth of 11 years across 83 companies to investigate the relationship between shareholder returns and different internal performance measures. A pooled panel regression model was employed to estimate common slope and intercept coefficients. Heterogeneity between firms was not controlled for and the study may have suffered from omitted variable bias.

Friss and Smit (2004) investigated the relationship between unit trust performance and risk measures and various unit trust-specific quantitative (e.g. fund age and manager years of education) and qualitative (e.g. whether or not managers possess CA/CFA qualifications) measures. They used a data set with a time depth of seven years across 57 unit trusts. Common slopes and intercepts were estimated, using a pooled regression model. Heterogeneity between unit trusts was not controlled for and the study may have suffered from omitted variable bias.

Heyns, Hamman and Smit (1999) investigated the relationship between share prices and future information about earnings in accruals and cash flows. They used a panel data set of 3 244 firm years over the period 1974 to 1996. According to Heyns *et al.* (1999:125), a single pooled regression was performed. Heterogeneity between firms was not controlled for and the study may have suffered from omitted variable bias. Similarly, in their study, Kruger, Steyn and Kearney (2002) did not control for possible heterogeneity between audit teams.

3.2.2 International journals

The Accounting Review

Lev (1989) reviewed market-based research on the information content of accounting earnings and found that the explanatory power of earnings for share returns was extremely low. His first possible explanation for the disappointing results was that the estimating equations were poorly specified because they did not allow for cross-sectional variation in the regression parameters.

Strong and Walker (1993) empirically examined Lev's first three explanations for the poor explanatory value of earnings for share returns and found that a significant improvement in the statistical performance of models of earnings and returns could be achieved by allowing for cross-sectional and time-series variation in the regression parameters. They did this by including an earnings yield variable as one of the explanatory variables and splitting earnings into pre-exceptional (standard) earnings, exceptional earnings and extraordinary earnings. The data panel included 146 companies over a time period of approximately 14 years. They employed a fixed effects approach in allowing for cross-sectional and intertemporal variation. Tests performed on the data included

descriptive statistics, correlation analysis to identify potential multicollinearity problems, and serial correlation tests. Potential shortcomings of the empirical work done were that no tests were performed on the suitability of the panel data approach adopted (a two-way fixed effects model specification was used) and the fact that there were no tests for heteroscedasticity.

Bhattacharya, Daouk and Welker (2003) used a data panel of 34 countries over 15 years to test the effect of opaque earnings on cost of capital and trading volume. Their study is remarkably complete in terms of the statistical techniques used to analyse and test the data panel. Another notable attribute of their study was the fact that they limited the heterogeneity of the cross-sections by confining their analysis to industrial firms in different countries only, so that differences in the underlying earnings process across different industry groups and differences in the proportion of firms in various industry groups did not affect the study. The study is unique compared to international studies on the value-relevance of accounting because extensive hypothesis testing was conducted (Bhattacharya *et al.* 2003:645). The study found that the potential challenges of omitted explanatory variables and endogeneity are mitigated by using panel data tests corrected for country-fixed effects, country-specific heteroscedasticity and country-specific autocorrelation. Their reasoning was that the use of panel data tests with fixed country effects would minimise the endogeneity problem. They specifically allowed for the endogeneity of variables by checking the validity of their inferences with a system of equations jointly estimated using SUR. The SUR specification computes estimates using the technique of joint GLS (generalized least squares).

Journal of Accounting and Economics

Banker, Chang and Cunningham (2003) used a balanced panel of data from 64 accounting firms over five years to estimate the relation between service revenue generated and human resources employed in the public accounting profession. Empirically they allowed for heterogeneity between firms and estimated a system of equations jointly using a SUR specification. This allowed them to vary intercepts and slopes per firm and to overcome any problems with endogeneity. Allowing slopes to vary per firm and over time was a specific consequence of their testing for an increase of economies of scale in the accounting profession.

An interesting study by Pittman and Fortin (2004) used data from 371 firms over the first nine years of public life to show that choosing a “Big Six” auditor enables young firms to lower their interest rates. They pruned their data by removing extreme observations from the dependent variable to remove noise from the data set. Their panel data model specification controlled for unobserved firm-specific effects to avoid omitted variable bias. They specifically used a two-way fixed effects model with correction for unspecified heteroscedasticity. Serial correlation was accounted for by checking for the robustness of the statistical evidence after correcting for serial correlation.

Journal of Accounting Research

Banker, Devaray, Schroeder and Sinha (2002) used data over 36 months for 18 plants to show that plants that stop monitoring direct labour variances experience a decrease in productivity and an increase in quality. Two separate fixed effects models were estimated (Banker *et al.* 2002:1020). Test results for serial correlation, heteroscedasticity and multicollinearity were reported and corrections made for serial correlation.

3.2.3 *Conclusions based on the literature review*

The South African literature review displayed little evidence that panel data techniques are utilised to their maximum potential. Only one example was found where allowance was made for heterogeneity in cross-sections.

In contrast, the international evidence showed that panel data techniques are well entrenched in the statistics arsenal of accounting researchers. Of note is the fact that most of the accounting studies that use panel data specific techniques are fairly recent; a possible conclusion is that panel data-specific techniques and their usage are still fairly new in accounting research⁴.

4 Empirical research

4.1 Objective

The objective of the empirical research is to illustrate the misleading inference that could result from the use of inappropriate estimation methodologies in panel settings. For this purpose, a model of the relationship between a company's market value and chosen accounting variables is used.

However, it is important to note that this model is used purely for illustrative purposes and that the economic implications of the findings are not discussed.

4.2 Methodology

Data from a previous study that investigated the relationship between five accounting measures and a company's market value are utilised as a reference and panel data techniques and tests applied to the data. Where a common intercept and common slopes were estimated, the regression results are then compared to the results of the panel data techniques allowing for heterogeneity across firms, and any significant differences are highlighted.

4.3 Data

The source of the data used in the study was McGregor BFA, based at the University of Pretoria. McGregor BFA is a major provider of information for the financial analysis of South African listed and de-listed companies. Their Station product is a fundamental analysis tool, which contains information on listed companies, de-listed companies, commodities, N-shares and preference shares. What makes their data especially useful for accounting research is that they capture the data contained in the annual financial statements of listed companies on an annual basis, standardise the data and calculate 42 standard financial ratios.

The final data set used in the study consists of a matrix of 13 years of observations of five financial ratios for 53 companies. Each component of this 13x5x53 matrix will now be discussed in greater detail.

⁴ The use of panel data techniques in other disciplines dates back to the early 1970s.

The years 1990 to 1994 can be regarded as a changeover period for South Africa from the political system of apartheid to one of a representative democracy. This political change would have influenced the South African capital markets. It was thus decided to limit the data used in the study to a period after this structural break. Capital markets are forward looking, and since the negotiations ushering in the post-apartheid South Africa were finalised at the beginning of 1993, this year was taken as the breakpoint. Data were therefore obtained for the 13 years from 1993 to 2005.

The five financial ratios obtained for this study are included in table 1.

Table 1 Variables used in the study

Dependent variable:	
MVA (market value added)	MVA is the increase in market value of a firm.
Independent variables:	
IC _{beg} (invested capital at the beginning of the year)	IC _{beg} is the capital that was used to generate economic profits as measured by the financial statements.
Sales growth	This is the growth in sales from the previous period.
Spread	Spread is the performance spread which is standardised economic value added (EVA/IC _{beg}). EVA is a performance measure calculated after taking into account the full cost of capital, including an opportunity cost for using equity.
Stdcashflow	Stdcashflow is cash flow from operations standardised by dividing by IC _{beg} .

All companies listed on the JSE on 25 January 2006 were identified as a first step in choosing the companies from which financial ratios would be obtained, a total of 325 companies. Banks, other financial institutions and mining companies were then excluded because they could not provide the required information to determine the critical variables for the analysis. ALT_x, Development Capital and Venture Capital listed companies were also excluded. This reduced the total to 177 companies. Thinly traded shares were identified and excluded from the study (113 in total). A further 11 companies were finally eliminated on account of missing data or financial year-ends that changed during the 13 years under review. This left 53 companies for which the five financial ratios were obtained for the 13 years from 1993 to 2005.

4.4 Results

The regression results for different model specifications are indicated below. Since hypothesis testing indicated that the data could not be pooled and that fixed effects were valid, estimation results allowing for individual fixed effects are also reported. Hypothesis testing indicated that heteroscedasticity and serial correlation are present in the error terms and estimation results with corrections for these problems are also reported.

Table 2 Selected regression results

X-regressors (excluding intercepts)							
No.	Model specification	Invested capital coefficient	Sales growth coefficient	Spread coefficient	Stdcashflow coefficient	Adjusted R squared	Probability (F-statistic)
1	Data pooled and common intercepts and slopes	0.3803 (0.0000)	1186.621 (0.8702)	55112.99 (0.0270)	22267.30 (0.0266)	0.1829	0.0000
Pooling & validity of fixed effects tests							
2	Common slopes with cross-section-specific intercepts	0.3273 (0.0000)	4040.795 (0.4962)	33526.00 (0.1311)	8212.141 (0.3329)	0.5165	0.0000
Serial correlation tests							
3	Common slopes with cross-section-specific intercepts, corrected for serial correlation	0.1473 (0.0079)	2596.941 (0.5863)	-9886.770 (0.6318)	3605.819 (0.5980)	0.2810	0.0000
Heteroscedasticity tests							
4	Common slopes with cross-section-specific intercepts, corrected for heteroscedasticity and serial correlation	-0.0004 (0.9967)	1043.140 (0.0019)	1752.899 (0.0079)	211.7454 (0.1484)	0.3763	0.0000

The relationship between accounting data and market data is often examined in the accounting literature. A relationship, but not a strong one, is expected between the explanatory accounting variables and MVA. Other factors besides accounting results also drive market prices, such as market perceptions and expectations, inflation, exchange rates and economic conditions. Hence omitted variable bias can be expected in the regression.

A positive relationship was expected for the explanatory variables. In some cases it could be argued that the market punishes the share price of a firm that wastes capital on projects that do not earn in excess of its cost of capital. An ambiguous result for IC_{beg} is therefore not surprising.

Multicollinearity between explanatory variables was investigated using a correlation table. The highest correlation was less than 0.40. Multicollinearity between the explanatory variables is therefore not considered to be a problem here.

The first step was to perform a pooled regression in which a common intercept term and common slope coefficients were estimated. The results labelled "no. 1" indicate that the explanatory power of the regression was extremely low (R^2 of 0.1829) with the coefficients of IC_{beg} , Spread and Stdcashflow statistically significant. The F-statistic of the regression indicates rejection of the H_0 of the coefficients of all explanatory variables simultaneously equal to zero.

A poolability test was performed to test the null of a common intercept and slope coefficients versus the alternative of running individual regressions for each cross-section. The calculated F-statistic ($F=4.7130$) exceeds the critical value of 1.0000 ($F_{crit}=F_{(N-1)K',(N(T-K'))};0.05$) and the null hypothesis of a common intercept and common slope coefficients for all cross-sections is rejected, and the data should therefore not be pooled for regression purposes.

The validity of fixed effects and fixed time effects was then tested. The calculated F-statistic ($F=8.7892$) (testing for pooled versus individual and time fixed effects) exceeds the critical value of 1.32 ($F_{crit}=F_{(N+T-2),((N-1)(T-1)-K)};0.05$), the null hypothesis is rejected and

fixed effects and time fixed effects are jointly valid. The calculated F-statistic ($F=10.0771$) (testing for pooled versus individual fixed effects) exceeds the critical value of 1.35 ($F_{crit}=F_{(N-1),(NT-N-K);0.05}$), the null hypothesis is rejected and individual fixed effects are valid. The calculated F-statistic ($F=1.1407$) (testing for pooled versus time fixed effects) did not exceed the critical value of 1.75 ($F_{crit}=F_{(T-1),(NT-T-K)-K;0.05}$) – hence the null hypothesis is not rejected and time fixed effects are not valid.

The results labelled “no. 2” indicate the results of the regression with allowance made for firm heterogeneity using fixed effects. R^2 has increased to 0.5165, implying that 51.65% of the variation in the dependent variable, MVA, can be explained by variation in the explanatory variables. Most of the slope coefficients have changed materially and the statistical significance of coefficients has changed in the case of Spread and Stdcashflow. The F-statistic of the regression indicates rejection of the H_0 of the coefficients of explanatory variables simultaneously equal to zero.

The results labelled “no. 2” are totally different from the pooled results labelled “no. 1”. This clearly illustrates that the results of accounting studies that use panel data and pool the data can be misleading. Further econometric tests deemed essential in this panel context are tests for serial correlation and heteroscedasticity.

A test for serial correlation given fixed effects showed that first-order serial correlation is a problem. The LM test has a value of 9.0467 and the 95% critical value is 1.96. The H_0 of no serial correlation is rejected. The results labelled “no. 3” were corrected for serial correlation by the Prais-Winsten transformation per cross-section⁵ and the LM test done after the correction had a value of 0.1761. This did not result in rejection of H_0 .

The results labelled “no. 3” indicate that the coefficients of IC_{beg} , Salesgrowth, Spread and Stdcashflow changed significantly. Statistical significance of the explanatory variables is still the same as in the results labelled “no. 2”. R^2 decreased to 0.2810 and the F-statistic of the regression indicates rejection of the H_0 of the coefficients of explanatory variables simultaneously equal to zero.

A visual inspection of the residuals of the different cross-sections reveals significant size differences in the residuals per cross-section. The LM statistic of a formal test for heteroscedasticity, $LM=1\ 311.2915$, exceeds the critical value of 43.773 (Chi-square(N-1) distributed), the null hypothesis of homoscedasticity is rejected and the alternative of heteroscedasticity in the error terms seems highly likely and will have to be corrected.

Heteroscedasticity was corrected for by using the White cross-section coefficient covariance method to correct the standard errors. This estimator is robust to cross-section (contemporaneous) correlation as well as different error variances in each cross-section. In addition, GLS with cross-section weights was used to allow for heteroscedasticity in the relevant dimension. This double correction for heteroscedasticity follows the study of Kyriazis and Anastassis (2007) who corrected for serious cross-sectional heteroscedasticity problems (the study in this paper also suffers from serious cross-sectional heteroscedasticity problems) by using weighted least squares and White’s correction for heteroscedasticity. The use of weighted least squares in addition to White’s correction also leads to a much better fit of the regression to the data. In comparison with the use of only White’s

⁵ Owing to the difference in the sizes of the cross-sections (firm), individual rho values for each cross-section were calculated.

correction, \bar{R}^2 increases to 0.3763 from 0.2810 and the standard error of the regression decreases from 5 247 455 to 4 734 845.

The final regression labelled “no. 4” is materially different from the regression labelled “no. 1”. The explanatory power of the regression (\bar{R}^2) has more than doubled, the slope coefficients are materially different, IC_{beg} is no longer statistically significant and Salesgrowth and Spread are now statistically significant. The F-statistic of the regression still indicates rejection of the H_0 of the explanatory coefficients simultaneously equal to zero. An omitted variable bias problem is likely to be less significant in this last regression. This final regression has survived a battery of statistical tests.

It was shown that the use of panel-specific statistical techniques can significantly improve and alter regression results in an accounting research setting.

5 Conclusion

Accounting researchers that utilise firm level data frequently build data sets with a time and a firm dimension. This panel of data has unique attributes that can improve the analysis, but specific econometric techniques are needed.

The literature review showed that South African accounting researchers often use panel data without using the econometric techniques available. This leads to statistical problems such as omitted variable bias and low degrees of power and this could even invalidate previous findings. In contrast, the international literature review shows widespread use of panel data techniques. Most of the international studies are recent and a possible implication is that panel data techniques also need to be adopted and applied in South Africa.

The empirical section showed that regression results can change significantly when panel data techniques are used. Accounting studies should allow for heterogeneity between firms, and the use of panel data techniques could assist in achieving this goal.

Bibliography

- Baltagi, B.H. 2005. *Econometric analysis of panel data*. 3rd edition. West Sussex: Wiley.
- Banker, R.D., Chang, H. & Cunningham, R. 2003. The public accounting industry production function. *Journal of Accounting and Economics*, 35:255-281.
- Banker, R.D., Devaray, S., Schroeder, R.G. & Sinha, K.K. 2002. Performance impact of the elimination of direct labor variance reporting: a field study. *Journal of Accounting Research*, 40(4):1013-1036.
- Bhattacharya, U., Daouk, H. & Welker, M. 2003. The world price of earnings opacity. *The Accounting Review*, 78(3):641-678.
- De Jager, P.G. & De Wet, J.H.v.H. 2007. An appropriate financial perspective for a balanced scorecard. *South African Business Review*, 11(2):98-113.
- De Wet, J.H.v.H. 2005. EVA versus traditional accounting measures of performance as drivers of shareholder value: a comparative analysis. *Meditari Accountancy Research*, 13(2):1-16.

- De Wet, J.H.v.H. & Du Toit, E. 2007. Return on equity: a popular, but flawed measure of corporate financial performance. *South African Business Review*, 11(1):1-38.
- Firer, S. & Stainbank, L. 2003. Testing the relationship between intellectual capital and a company's performance: evidence from South Africa. *Meditari Accountancy Research*, 11:25-44.
- Friss, L.B. & Smit, E.v.d.M. 2004. Are some fund managers better than others? Manager characteristics and fund performance. *South African Journal of Business Management*, 35(3):31-40.
- Gouws, D.G. & Lucouw, P. 2000. A dynamic balance model for analysts and managers. *Meditari Accountancy Research*, 8:25-45.
- Greene, W. 2008. *Econometric analysis*. 6th edition. Upper Saddle River, N.J.: Pearson/Prentice Hall.
- Heyns, P.J., Hamman, W.D. & Smit, E, v.d.M. 1999. Do share prices reflect the information about future earnings in accruals and cash flow? *South African Journal of Business Management*, 30(4):122-130.
- Kruger, H.A., Steyn, P.J. & Kearney, W. 2002. Determinants of internal audit efficiency. *South African Journal of Business Management*, 33(3):53-61.
- Kyriazis, D. & Anastassis, C. 2007. The validity of the economic value added approach. *European Financial Management*, 13(1):71-100.
- Lev, B. 1989. On the usefulness of earnings and earnings research: lessons and directions from two decades of empirical research. *Journal of Accounting Research*, 27(supplement):153-201.
- Lind, D.A., Marchal, W.G. & Wathen S.A. 2005. *Statistical techniques in business and economics*. 12th edition. New York: Mcgraw-Hill.
- Negash, M. 2001. Uncertainty, cost of capital and financial disclosure. *South African Journal of Accounting Research*, 15(2):49-76.
- Pittman, J. & Fortin, S. 2004. Auditor choice and the cost of debt capital for newly public firms. *Journal of Accounting and Economics*, 37(1):113-136.
- Steyn, B.W. & Hamman, W.D. 2004. The time series behaviour of net profit, cash flow from operating activities and accruals of South African listed industrial companies for the period December 1988 to November 2002. *South African Journal of Accounting Research*, 18(1):115-131
- Strong, N. & Walker, M. 1993. The explanatory power of earnings for stock returns. *The Accounting Review*, 68(2):385-399.
- Swartz, N.P. & Firer, S. 2005. Board structure and intellectual capital performance in South Africa. *Meditari Accountancy Research*, 13(2):145-166.
- Swartz, G.E., Swartz, N.P. & Firer, S. 2006. An empirical examination of the value relevance of intellectual capital using the Ohlson (1995) valuation model. *Meditari Accountancy Research*, 14(2):67-81.
- Van Staden, C.J. 1999. Aspects of the predictive and explanatory power of value added information in South Africa. *South African Journal of Accounting Research*, 13(2): 53-75.