Analysts' Forecasts as Rational Expectations of Company Earnings

Geoffrey Brooke

2004

Submitted in fulfilment of the requirements for the Degree of Master of Business Science.

Abstract

This study is a test of the hypothesis that analysts' forecasts are rational expectations of earnings for firms listed on the JSE Securities Exchange. The study covers the period January 1990 to December 1999.

To be considered rational expectations the forecasts need to be (i) the most accurate available, (ii) unbiased and (iii) have no exploitable patterns in the errors.

Weak evidence of an optimistic bias is found in the mean errors. The magnitude of the error is found to be inversely related to firm size and the magnitude of the error in the previous period. The bias is found to be inversely related to the bias in the previous period.

Analyst forecasts are more accurate than forecasts produced by nine models based on extrapolations of past earnings. Two time-series models produced less biased forecasts.

Analyst forecasts do not meet the strict requirements of the Rational Expectations Hypothesis.

Contents

List of Tables

Chapter 1

Introduction

The value of a firm is a function of the expected future cash flows and the riskiness of the cash flows, or the cost of capital. In estimating future earnings, stock analysts solve the first part of the security valuation problem.

This thesis tests the hypothesis that analysts' forecasts are rational expectations of firm earnings. Given the above, the question is central to the usefulness of analysts' forecasts for the purposes of security valuation.

Following the criteria of Muth (1960,1961), in order to be considered rational expectations the forecasts need to meet three requirements, (i) the mean forecast error must be zero, (ii) they must be the most accurate forecasts available and (iii) there must be no patterns in the errors.

The analysis is based on consensus earnings forecasts for firms listed on the JSE Securities Exchange. The forecasts cover the period from January 1990 to December 1999.

The study is organised as follows. The theoretical basis for the study is discussed in Chapter Two. The discussion is divided into two parts; the first is a derivation of the earnings discount valuation model with a brief discussion of the related empirical evidence; the second part is a discussion of the Rational Expectations Hypothesis.

Chapter Three describes all data used in the study.

Each of the requirements for rational expectations is then empirically tested in turn. Given the largely independent literature covering the three questions into which the null hypothesis is divided, the discussion of prior research is, in each case, contained in the relevant chapter where the question is considered.

Chapter Four consider the question of bias in analysts' forecasts. The accuracy of analysts' forecasts are compared to alternative forecasts for firm earnings in Chapter Five. Specifically, analyst forecasts are compared to those generated from econometric models based on the time-series of past earnings.

Chapters Six and Seven consider the question of whether the errors that the analysts make are predictable. These chapters differ primarily in the nature of the sample used. A significant innovation in this work is the methodology employed in testing the predictability of forecast errors. Whereas previous studies have used ordinary least squares regression, this study is based on a fixed effects panel approach, which produces more robust results. Chapter Eight concludes the study.

Chapter 2

Theoretical Overview

2.1 Introduction

This chapter contains a discussion of the theoretical importance of earnings. This is followed by a brief discussion of the empirical findings on the relationship between earnings and share prices, or more specifically, unexpected changes in earnings and share returns. The chapter concludes with a discussion of rational expectations and the importance of rational expectations in analyst forecasts.

Keane and Runkle(1998) offer three reasons for the particular interest in analysts' forecasts, and the question of whether they are rational. First, asset pricing models generally require expected earnings as an input. While forecasts based on time-series models are available for and are used as a proxy for market expectations, they create two problems. First, they do not use the full set of information available to the forecaster. Second, testing an asset pricing model where expectations are based on time-series forecasting models involves a joint test of the forecasting model and the asset pricing model.

Their second reason is that if analysts' forecasts are not rational, then analysts' forecasts may be a part cause of what,'some researchers consider are excessively volatile asset price movements'(p769). The final reason is that analysts forecasts provide an opportunity for direct testing of the rational expectations hypothesis. Given that,'financial analysts' livelihoods depend on the accuracy of their forecasts and since we observe the same forecasts that analysts sell, we can plausibly argue that these numbers accurately measure the analysts' expectations'(p769).

2.2 Earnings and the value of the firm

The value of any ordinary share can be expressed as a function of the current earnings, the expected growth in earnings and an appropriate discount rate. The equivalence of the capitalization of earnings and dividends was originally demonstrated by Miller and Modigliani (1961), more recent treatment of earnings discount models can be found in Fama and Miller (1972) and Copeland and Weston (1988). A complete review of asset pricing and the related literature can be found in Campbell (2000). This discussion is summarised from Miller and Modigliani (1961), the notation follows theirs.

In addition to assuming perfect capital markets and rationality on the part of investors, it is assumed that there is perfect certainty about returns. The first portion of the derivation shown here demonstrates the irrelevance of dividends in valuing a firm. The second portion is the derivation of an earnings discount model that is identical to the final form of the dividend discount model. The purpose is to demonstrate the equivalence of earnings and dividend discount models in valuing a firm.

The holding period return to an owner of ordinary equity is the sum of the capital gains and any dividends paid during the holding period.

If

$d_j(t)$ = dividends per share paid by firm j during period t, and

$p_j(t)$ = the price of a share in firm j at the start of period t, then

$$\frac{d_j + p_j(t+1) - p_j(t)}{p_j(t)} = \rho(t). \tag{2.1}$$

Where $\rho(t)$ is the certain rate of return. The certainty assumption has two implications. First, the separation of debt and equity claims is removed from the analysis and second, risk is removed from the discount rate. The assumption can be relaxed either by assuming an all-equity funded firm or by appealing to Modigliani and Miller (1958) and assuming firms of an identical risk class and level of leverage, and treating $\rho(t)$ as the discount rate appropriate for a firm of that risk class and with the given leverage. A more modern approach would be to consider $\rho(t)$ to be the expected return for firms with the same systematic risk per the Capital Asset Pricing Model (CAPM).

Then, if

$n(t)$ = the number of shares in issue at the start of t

$m(t+1)$ = the number of new shares issued during t at the ex dividend closing price $p(t+1)$,

$n(t+1) = n(t) + m(t+1)$,

$V(t) = n(t)p(t)$ = the total value of the company, and

$D(t) = n(t)d(t)$ = the total dividends paid to owners of the shares in issue at the start of t,

(2.1) can be rewritten as

$$V(t) = \frac{1}{1+\rho(t)}[D(t) + V(t+1) - m(t+1)p(t+1)]. \tag{2.2}$$

If $I(t)$ is the company's investment in physical assets in t and $X(t)$ is the firm's net profit for period t then the number of new shares issued will be

$$m(t+1)p(t+1) = I(t) - [X(t) - D(t)]. \tag{2.3}$$

Substituting (2.3) into (2.2) the value of the company at the start of t can be expressed as

$$V(t) \equiv n(t)p(t) = \frac{1}{1+\rho(t)}[X(t) - I(t) + V(t+1)]. \tag{2.4}$$

Having demonstrated that the value of the firm is independent of dividends paid during t, the argument can be extended to the value of the firm at $t+1$ (and thus the value at t) being independent of dividends paid during period $t+1$. Extending this argument, the value of the firm can then be stated as

$$V(0) = \sum_{t=0}^{\infty} \frac{1}{(1+\rho(t))^{(t+1)}} \times [X(t) - I(t)]. \tag{2.5}$$

The equivalence of discounting earnings and dividends can then be demonstrated. All future earnings cannot simply be discounted. The cost of the firm's future capital requirements, $I(t)$ with the required return of ρ percent per period, need to be taken into account. The current value of the firm can then be expressed as

$$V(0) = \sum_{t=0}^{\infty} \frac{1}{(1+\rho(t))^{(t+1)}} \times [X(t) - \sum_{\tau=0}^{t} \rho I(\tau)]. \tag{2.6}$$

This can be rewritten as

$$V(0) = \sum_{t=0}^{\infty} \frac{1}{(1+\rho(t))^{(t+1)}} X(t) - \sum_{t=0}^{\infty} \left(\sum_{\tau=t}^{\infty} \frac{\rho I(t))}{(1+\rho)^{\tau+1}} \right), \tag{2.7}$$

$$V(0) = \sum_{t=0}^{\infty} \frac{1}{(1+\rho(t))^{(t+1)}} X(t) - \sum_{t=0}^{\infty} \frac{1}{(1+\rho)^{t+1}} \times \left(\sum_{\tau=0}^{\infty} \frac{\rho I(t)}{(1+\rho)^{\tau+1}} \right). \tag{2.8}$$

The last summation is $I(t)$, and (2.8) reduces to

$$V(0) = \sum_{t=0}^{\infty} \frac{1}{(1+\rho(t))^{(t+1)}} \times [X(t) - I(t)]. \tag{2.9}$$

which is the same as (2.5).

To the extent that analysts are able to forecast future earnings, they solve one half of the share valuation problem.

The theoretical evidence on the importance of earnings is substantiated in the literature that reports empirical findings in line with expectations. Among the best known of these is Niederhoffer and Regan (1972). They find that the most significant difference between the fifty best performing and fifty worst performing stocks in 1970 was the change in earnings, and particularly the change in earnings relative to expectations. The worst fifty performing stocks had a median forecast change in earnings of 15.3% and a realized change in earnings of -83%. In contrast the top fifty performing stocks had a median forecast earnings change of 7.7%

and a realized median earnings change of 21.4%. These results demonstrate a clear link between earnings surprises and share returns.

Two studies have used forecast errors as an explanation for abnormal returns after earnings are announced. Foster, Olsen and Shevlin (1984) find that earnings surprises (based on time-series forecasting methods for determining expected earnings) explain 81% of post earnings announcement drift. They note that the, 'more positive (negative) the unexpected earnings change, the more positive (negative) the post-announcement abnormal returns'(p598). More recently Kim and Kim (2003) added an earnings surprise factor to the factors proposed by Fama and French (1993) for explaining stock returns. After the addition of the earnings surprise factor to the model the abnormal returns following the announcement of earnings are statistically indistinguishable from zero.

Perhaps an appropriate end to this discussion is a comment on the Ball and Brown (1968) study that provided the seminal work for the papers that constitute the discussion. In a response paper Brown (1989) states that the early paper was significant because, 'It documented an association between earnings and prices that has proved robust over time and different markets'(p205).

2.3 The Rational Expectations Hypothesis

Having established a theoretical relationship between share prices and expectations of future earnings that is supported by empirical evidence, it is necessary to discuss the characteristics of the forecasts. This includes prior expectations about the characteristics of analysts' forecasts, their implications for market efficiency and their usefulness in asset pricing as measures of expectations.

The Rational Expectations Hypothesis, Muth (1960,1961), is essentially an argument for the complete use of available information. The relationship between efficient markets hypothesis and the rational expectations hypothesis is highlighted in the following from Fama (1970:383),'A market in which prices always 'fully reflect available information is called 'efficient". Similarly, from Kantor (1979:1430), 'A proof of efficient use of information in markets is a proof of rational expectations'.

Linking the efficient markets and asset pricing problems is Easley and Jarrow (1983:911) who demonstrate that, 'unless traders have rational expectations, the concept of consensus beliefs is not useful for considerations of market efficiency'. Thus for analyst forecasts to be of value for the purposes of security valuation, the forecasts must be rational.

If the underlying process by which earnings are generated is stochastic, then the earnings announced at time t are made up of two components, a predictable portion, based on those factors that drive the underlying process, and a random component that cannot be forecast with information known before earnings are announced.

The process could then be described in the following way:

$$A_{it} = \alpha_i + \sum_{k=1}^{K} \beta_i \mathbf{X}_i + \epsilon_{it}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T. \tag{2.10}$$

Where A_{it} is the actual earnings for firm i at time t, α and β are 1×1 and $1 \times k$ vectors of coefficients, $\mathbf{X}$ is the $k \times 1$ vector of exogenous variables, and ϵ is the error term that is assumed to be identically and independently distributed with zero mean and a variance σ^2.

Given an *a priori* expectation that analysts' forecasts are rational then, from Attfield, Demery and Duck (1985), they are expected to have the following characteristics:

1. The forecasts are unbiased.
2. There are no discernable patterns in the errors.
3. The forecasts are the most accurate forecasts possible.

If the forecast bias is not zero, then the error term ϵ will have a constant component. A rational market participant will recognise this and produce forecasts that include this information in the intercept term. Thus the mean error for rational forecasts may not be either positive or negative. This is not to say that individual observations will all have a zero error, simply that on average the error is zero. A problem, identified by Crichfield, Dyckman and Lakonishok (1979), in testing the rational expectations hypothesis is that system shocks (by definition impossible to forecast) tend to have effects that are correlated in the cross section that may give the appearance of biased forecasts if short time periods are used in the analysis.

The argument concerning patterns in the errors is similar. If the error term contains a variable but forecastable component (is itself a stochastic process), then the rational forecaster will form an expectation of the value that the forecastable component will take in the next period. At the same time a forecast of zero will be maintained for the strictly random part. There will then be no patterns in the errors that forecasters make if their forecasts represent rational expectations.

As an argument on the use of complete information, if the error term, or a part of it, can be forecast from information available when the forecast is made, then the forecast cannot represent rational expectations for firm earnings. From Lovell (1986:113), 'The prediction error must be uncorrelated with the entire set of information that is available to the forecaster at the time the prediction is made'. Similarly, from Kantor (1979:1430), 'It is however incorrect to assume that rational expectations regards errors in forecasts as insignificant or absent. The implication of rational expectations is rather that the forecast errors are not correlated anything that could be profitably known when the forecast is made.'

Finally, if analysts' forecasts reflect all available information of the underlying process by which earnings are generated then, given the fact that the error term is on average zero, it must not be possible on average to produce a forecast more accurate than the analysts.

Criticisms of the Rational Expectations Hypothesis

There are a number of standard criticisms (Attfield et al (1985)) of the rational expectations hypothesis, some of which are answered by the nature of the data used in this study. The first question, similar to the attacks on utility theory is simply whether it is ever reasonable to assume that people hold rational expectations. It is certainly not necessary for all agents in the economy to acquire and analyse data in order to form rational expectations in order for the rational expectations hypothesis to hold. All that is required is that those who have an economic interest in doing so either analyse the data themselves or purchase forecasts from those who produce forecasts with the express purpose of selling them and whose livelihood depends on accurate forecasts.

The second criticism is similar in that it introduces a cost to information and asks whether it is reasonable to assume rational expectations if forecasters collect information only to the point where the marginal value of the information is equal to the marginal cost of information. This criticism is perhaps more relevant, although the response is similar. As long as there are participants in the economy whose livelihood depends on the accuracy of their forecasts, then it appears to be reasonable to assume that relevant information about the underlying process by which earnings are generated is included in the forecasts.

A third criticism is that there is no recurring stochastic process underlying the earnings generation process. This criticism has been confused with shocks (see Attfield et al (1985:27) and the discussion of British monetary policy around the Falklands War). Shocks are, as discussed, a problem in empirical analysis. They are not a problem for the theory of rational expectations. As to whether there is an underlying stochastic process, the simplest description of the firm is that given in (2.5) above. The expected future earnings are determined by the riskiness of the assets and the amount of money invested in those assets. It does therefore appear reasonable to assume that there is a stochastic process underlying earnings generation.

A final common criticism that is not relevant here is that empirical testing of forecasts has typically involved forecasts based on a time-series forecasting method that results in a joint test of the rational expectations hypothesis and the forecasting model. As per Keane and Runkle (1998), the benefit of working with analyst forecasts is that they are made in advance, and if they are a reasonable reflection of market expectations, the rational expectations hypothesis can then be tested directly.

Chapter 3

Data

3.1 Introduction

All data used in the empirical analysis are described in this chapter. The use and reasons for choice of data are explained in the appropriate chapter.

3.2 Earnings and earnings forecast data

Definitions of the analyst forecast error are based on the consensus earnings forecast of financial analysts for full year earnings and the realised full year earnings.

The consensus earnings forecasts are taken from the Institutional Broker Estimation Service (I/B/E/S) database. The database used consists of monthly earnings forecasts for full-year earnings per share for the period January 1990 to December 1999. The I/B/E/S consensus forecast is the median of individual analyst forecasts after the removal of all forecasts lying more than 2.3 standard deviations from the mean.

Actual full year earnings are from I-Net Bridge. Where appropriate the earnings are adjusted for stock splits. The I-Net Bridge earnings figures are full year headline earnings per share. Where the reporting period is less than one year,

Table 3.1: **Sample by Year and By Industry** – The consensus forecast for the month immediately prior to the earnings announcement is paired with the realised earnings.

	Industrials	Financials	Resources	Full Sample
1990	34		9	43
1991	39		11	50
1992	40		11	51
1993	40		11	51
1994	43		11	54
1995	45	2	11	58
1996	52	3	14	69
1997	56	7	15	78
1998	63	19	16	98
1999	75	21	17	113
	487	52	126	665

the figure is annualized, with no seasonal adjustment. For goldmines the earnings quoted are 'after-capex'.

The data can be divided into financial, resources and industrial firms. While the three industry split has limitations, the size of the database makes a finer split impractical. Brown (1997b) limits his study to industries with at least fifty firms for which data is available, leaving fourteen industries. By that standard, the industrial grouping used would only be eligible from 1996, and the resource and financial groupings would be omitted entirely. The industry subgroups are referred to as Industrial, Resources and Financials for the remainder of the thesis. This is done to distinguish subgroups within the sample from all listed stocks that would fall into each of the groups.

The number of companies for which earnings are forecast is 43 in 1990 and 113 in 1999. The total number of paired forecast and actual earnings is 665. The number of observations available by industry grouping and year is presented in Table 3.1.

3.3 Other data

The remaining data discussed are used in Chapters Six and Seven. In all circumstances the value used is the closing value for the month before earnings are announced.

Beta (BETA)

The beta used is calculated from 60 months of historical returns (Copeland and Weston, 1988, p214) where all price data are taken from I-Net Bridge. The beta is calculated using returns on the All Share Index (ALSI). ALSI index values are taken from I-Net Bridge.

Firm Size (LNMC)

Firm size data are taken from I-Net Bridge, where data are typically available from 1993. The natural log of the market capitalization is used as a measure of firm size.

Earnings Yield (EY)

Earnings yield is taken directly from I-Net Bridge. Data are available for the full ten-year period of the study.

Dividend Yield (DY)

Dividend yield is taken directly from I-Net Bridge. Data are available for the full ten-year period of the study.

Historical Growth (HGRO)

Historical growth is the arithmetic five year mean earnings growth. It is calculated from historical earnings data taken from I-Net Bridge.

Table 3.2: **Explanatory Variables**

Explanatory Variable	Code
Earnings Yield	EY
Dividend Yield	DY
Market Capitalization	LNMC
Historical Earnings Growth	HGRO
Volatility of Past Earnings Growth	HVOL
Inflation	INF
Interest Rate	PRIME
Beta	BETA
Lagged Absolute Percentage Error	ABSPER_L
Lagged Signed Percentage Error	PERA_L

Historical Earnings Volatility (HVOL)

Historical earnings (growth) volatility is the standard deviation of the changes in earnings over the past five years.

Inflation (INF)

Inflation is the 12 month change in the consumer price index (major metropolitan areas, all items), taken from I-Net Bridge.

Interest Rate (PRIME)

The mean 12 month prior prime interest rate is taken from I-Net Bridge.

Lagged Forecast Error (ABSPER_L, PERA_L)

The lagged forecast error is the error of the appropriate measure made in forecasting earnings in the previous period.

The forecasts used are short horizon forecasts, and thus do not provide a complete set of future earnings as are required for security valuation. They do however represent a potentially useful input as they provide an implied growth rate for future firm earnings.

It is important to note two limitations of this study that arise from the data used. First, the companies do not all have the same year end. The market conditions faced by an analyst forecasting for a January year end may be different from those for an analyst forecasting for December of the same year. Given the sample size and the lack of uniformity it is not feasible to work with a single year end. Second, there is often a lag in the updating of the I/B/E/S consensus forecasts, and the length of this lag is variable. In order to ensure that there is no overlap in reporting periods, the actual earnings are compared to the recorded consensus forecast from the previous month. Actual announcement months are taken from I-Net Bridge. The forecast used may, depending on the lag, be the most recent forecast, one, two or even three months old. As forecasts improve as the year progresses, the accuracy of analysts' forecasts may be understated. On the other side of the reporting date, the earnings forecasts for 12, 11 and 10 months before next year's earnings announcement may be for the previous year's earnings. There is no way of consistently correcting for this error as the lag is variable and does not appear to affect all stocks.

Chapter 4

Bias in Analysts' Forecasts [1]

4.1 Introduction

This chapter attempts to answer the first question posed; is the mean forecast error significantly different from zero? The chapter continues with a discussion of prior research. This is followed by the summary statistics relating to the forecast errors made by analysts, and compares them to international findings. Also included is a preliminary analysis on the source of the errors.

4.2 Prior research

The earliest evidence on forecast accuracy in the 1960s suggested a pessimistic bias in analyst forecasts. From the mid-1970s to the mid-1980s the evidence was mixed. Since the mid-1980s the empirical evidence is largely of an optimistic bias in analysts forecasts. The evidence up to the mid-1980s was limited to research conducted using data from the American financial markets. Since then research

[1] The results contained in this chapter were first presented at the Southern African Finance Association annual conference in January 2003. An earlier draft of this chapter was presented at the biennial Conference of the Economic Society of South Africa in September 2003. The full text of that paper is available at http://www.essa.org.za/download/papers2003.htm.

has expanded to include data from Europe, Asia and Australia.

Pre-1985 research

In an early study Cragg and Malkiel (1968) examine the accuracy of five year earnings growth forecasts from five sources (institutions involved in money management in some form) made at the end of 1962 and 1963. They find that analysts underestimated earnings growth, 'a result of the average growth rates being considerably higher than the average expectation of each predictor'(p77).

The first direct test of published forecasts came from MacDonald (1973) who analysed the accuracy of forecasts made by firms for their own earnings. The forecasts were published in the Wall Street Journal in time to be included in the annual financial statements, where the financial statements are published within 120 days of the financial year end, and the forecasts are for the following financial year. The forecasts were thus published eight to twelve months before the financial year end. For a sample of 201 observations spread over the 1966–1970 period the signed percentage error (actual earnings less forecast earnings, scaled by actual earnings) is –13.6%, with a standard deviation of 41.6%. This result indicates an optimistic bias in the forecasts. The null hypothesis that optimistic and pessimistic forecasts occur at the same frequency is rejected at the 5% level for the full sample and all years except 1966, confirming the optimistic bias in the forecasts.

Richards (1976) finds a small pessimistic bias analyzing forecasts from each of five analysts for 1972 earnings for a sample of 93 NYSE listed stocks. He finds significant differences in the errors between different industries but not between individual forecasters. Ultimately he concludes that, 'analysts perform reasonably well in forecasting earnings'(p357).

For the period 1972–1976 Richards, Benjamin and Strawser (1977) find an overall pessimistic bias in mean analyst forecasts taken from the Standard and

Poor's Earnings Forecaster. The bias is however optimistic in three of the five years of the study. Differences in errors between industries were again found to be significant.

Using Standard and Poor's Earnings Forecaster data from 1967–1976 Crichfield, Dyckman and Lakonishok (1978) regress the natural log of the actual change in earnings on the natural log of the forecast change and find intercepts that are not significantly different from zero. They do however note, 'a tendency for α to be negative on average'(p662), indicating weak evidence of an optimistic bias. Their slope coefficients are consistently greater than one, although not significantly different. A 1% forecast change corresponding with a realized change of greater than 1%. In addition they find that analysts' forecasts improve as the time to earnings announcement approaches. Elton, Gruber and Gultekin (1984) find similar results using I/B/E/S data from 1976–1978; the absolute percentage error decreases as the announcement date for earnings approaches with errors less than 5% in the month before earnings announcement. They do not provide any information on the direction (bias) of the analysts' errors.

Brown, Foster and Noreen (1985) use the I/B/E/S database for January 1976–December 1980 and the Wells Fargo database of individual analysts forecasts for January 1977–June 1980 to analyze the properties of analysts' earnings forecasts. Consistent with prior and subsequent evidence they find that the forecast error, regardless of definition, declines as the announcement date for earnings approaches. By number of forecasts they find that analysts are more likely to underestimate than overestimate earnings, and the median forecast is pessimistic. The mean signed percentage errors are however optimistic in all months except the one immediately preceding earnings announcement, suggesting that the magnitude of optimistic errors is greater than the magnitude of pessimistic errors.

Post-1985 research

The evidence to this point indicates either a pessimistic bias on the part of analysts, or is mixed as in the case of Crichfield *et al* (1978) and Brown *et al* (1985). The evidence to follow typically reports a more recent optimistic bias on the part of analysts, and this finding is repeated around the world.

De Bondt and Thaler (1990) investigate bias in analysts' forecasts for the period 1976–1984 using consensus I/B/E/S forecasts. Regressing the actual change in cents per share on the forecast change they find a significantly negative intercept, with a slope less than one, again indicating an optimistic bias.

Dreman and Berry (1995) use the Abel Nosser database of consensus earnings forecasts from 1972–1991. They find a significant optimistic bias in the mean forecast for the full sample. Further evidence of the optimistic bias is the number of optimistic and pessimistic errors - they found 26,122 positive surprises and 29,363 negative surprises in a sample of 66,100 observations. Dividing their sample into expansionary and recessionary periods for the economy as a whole, they find that the business cycle does not contribute in a significant way to either the size or the direction of analysts' errors. Finding that fewer than 30% of all forecasts fell within a 5% error bandwidth around the actual earnings, they concluded that the size of the forecast errors are, '...too high for investors to rely on consensus forecasts as a major determinant of stock valuation'(p 39).

Brown (1996) argues against a number of Dreman and Berry's (1995) findings, particularly their finding that the mean error is increasing through time. In a more complete paper Brown (1997b) finds similar results to Dreman and Berry (1995) - significant optimistic bias - using the I/B/E/S data for the period 1983–1996. He however finds smaller errors and bias for firms in the S&P500 than for those not in the index, and that no bias is observable for the S&P500 for the 1993–1996 period. While Dreman and Berry (1995) argue that forecast errors have increased through time, Brown (1997b) again finds the opposite. In Brown's sample the last year in which the annual mean bias exceeds the overall mean bias is 1991.

Using Value Line forecasts, important because Value Line is not involved in any broking or underwriting activities, Das, Levine and Sivaramakrishnan (1998) report a mean signed error of -1.5% at a three month forecast horizon. At this horizon 42% of the forecasts are optimistic.

Recent evidence from the European markets comes from Capstaff, Paudyal and Rees (2001). Using individual analysts' forecasts taken from I/B/E/S their study covers the period from February 1987 to September 1995 for nine countries; Belgium, France, Germany, Ireland, Italy, Netherlands, Spain, Switzerland and the United Kingdom. They find an optimistic bias in each of the nine countries in the percentage errors. Testing the bias more formally they regress the actual percentage change on the forecast percentage change they find significantly negative intercepts for all of the countries except Ireland where the intercept is significantly positive. While these findings on the intercepts are consistent with the findings of Crichfield *et al* (1978), like De Bondt and Thaler (1990) their slope coefficients are consistently less than one.

Higgins (1998) analyses I/B/E/S data for the U.S., Japan and five European countries for the period 1991–1995. The European countries are the United Kingdom, Netherlands, France, Germany and Switzerland. An optimistic bias is found in the percentage errors of all countries in the sample. The errors for the European countries are larger than those found by Capstaff *et al* (2001) for a similar period.

Further evidence on Japanese forecasts comes from a number of sources. Perhaps the most interesting is Conroy and Harris (1995) who analyze both individual I/B/E/S forecasts from (primarily) sell side analysts and Toyo Keizai forecasts for the period 1988–1992. Toyo Keizai is a data provider that does not engage in any underwriting activities. They find an optimistic bias in both sets of forecasts, however the I/B/E/S forecasts were more optimistic. This difference is attributed to the incentives facing sell side analysts working in firms with underwriting relationships with the firm that they were analyzing.

Higgins (2002) considers the relative performance of American and Japanese analysts forecasting earnings for Japanese companies. The American analysts performed better, with a mean absolute percentage error of 28% (1989–1998) against the 41% (1992–1998) mean percentage error for the Japanese analysts. The American analysts did a better job of forecasting earnings for Japanese firms than for American firms, although the errors made by American analysts forecasting earnings for American firms were found to decline through time. This was found to be the result of a decline in the bias in the forecasts. This corresponds with the findings of Brown (1997b). Increasing forecast accuracy was not found in the forecasts for Japanese companies from either set of forecasters.

Allen, Cho and Jung (1997) analyze I/B/E/S forecasts for Japan, the U.S., Korea, Malaysia, the Philippines, Taiwan, Thailand, Hong Kong and Singapore for 1989, 1990, and 1991. They find optimistic forecast bias in the percentage errors for all countries in the sample. They find a larger bias and lower forecast accuracy in the U.S. than in Japan. Their sample period is not long enough to comment on the changes in forecast accuracy over time. Like Elton *et al* (1984) they find that the accuracy of forecasts improves as the announcement date of earnings approaches.

Summary

The recent international evidence suggests an optimistic bias in analysts' forecasts. The errors in the American market appear to be larger on average than in Europe or the (non U.S.) Pacific Rim. These errors are however found to be much smaller for S&P500 companies. A possible explanation is that the American results are less accurate simply because relatively smaller companies are covered in the U.S. than in other countries. In the two instances where the forecasts were produced by firms not involved in broking or underwriting the errors were smaller than typically found in studies using the I/B/E/S database.

A summary of the findings on the bias and magnitude of the analysts' forecast errors is presented in Table 4.1.

4.3 Methodology and results

4.3.1 Size and pattern of the errors

In order to test forecast accuracy the earnings announced in month t are compared to the consensus forecast published at the end of month $t-1$. There is no single best definition of forecast error. Four definitions are used in this study (the naming conventions SURPE and SURPF are taken from Dreman and Berry (1995)):

$$ABSURP_{it} = |A_{it} - F_{i,t-1}|, \tag{4.1}$$

$$SURPE_{it} = \frac{A_{it} - F_{i,t-1}}{|A_{i,t-12}|}, \tag{4.2}$$

$$SURPF_{it} = \frac{A_{it} - F_{i,t-1}}{|F_{i,t-1}|}, \tag{4.3}$$

$$ABSPER_{it} = \left|\frac{A_{it} - F_{i,t-1}}{A_{i,t-12}}\right|. \tag{4.4}$$

A_{it} is the actual earnings at the end of month t for firm i and $F_{i,t-1}$ is the forecast captured at the end of the month before the month in which earnings are announced. ABSURP is the absolute surprise in cents per share. It offers no guidance as to the direction of the error, and is biased by the level of the

Table 4.1: Prior international evidence on forecast bias and mean error–Where $SURPE = (ActualEarnings - ForecastEarnings)/|ActualEarnings|$ and $ABSPER = |((Actualearning - ForecastEarnings)/ActualEarnings)|$.

Country	Period	Mean SURPE	Mean ABSPER	Forecast Type and Source	Study
U.S.	1972	-1.6% to -2.6%	8.5% to 9.3%	Five Individual Sources	Richards (1976)
	1972–1991	-12.8%	24.1%	S&P Earnings Forecaster	Richards *et al* (1977)
	1972–1976	-25.0%	43.8%	Abel Nosser	Dreman and Berry (1995)
	1983–1996	-41.4%	91.6%	I/B/E/S	Brown (1997)
	1983–1996	-12.9%	41.8%	I/B/E/S - S&P500 only	"
	1989–1991	-11.2%	28.8%	I/B/E/S	Allen, *et al* (1997)
	1991–1995	-29.0%	45.0%	I/B/E/S	Higgins (1998)
	1989–1993	-3.0%	Not Reported	Value Line	Das *et al* (1998)
	1989–1998	Not Reported	35.0%	I/B/E/S US Analysts	Higgins (2002)
United Kingdom	1991–1995	-35.0%	56.0%	I/B/E/S	Higgins (1998)
	1987–1995	-6.5%	14.9%	I/B/E/S	Capstaff *et al* (2001)
Germany	1991–1995	-24.0%	38.0%	I/B/E/S	Higgins (1998)
	1987–1995	-6.6%	20.9%	I/B/E/S	Capstaff *et al* (2001)
France	1991–1995	-16.0%	50.0%	I/B/E/S	Higgins (1998)
	1987–1995	-10.1%	18.1%	I/B/E/S	Capstaff *et al* (2001)
Netherlands	1991–1995	-21.0%	44.0%	I/B/E/S	Higgins (1998)
	1987–1995	-4.7%	13.1%	I/B/E/S	Capstaff *et al* (2001)
Switzerland	1991–1995	-55.0%	72.0%	I/B/E/S	Higgins (1998)
	1987–1995	-9.5%	20.6%	I/B/E/S	Capstaff *et al* (2001)
Ireland	1987–1995	-2.4%	17.4%	"	"
Italy	1987–1995	-12.3%	25.8%	"	"
Belgium	1987–1995	-5.8%	18.5%	"	"
Spain	1987–1995	-10.7%	20.0%	"	"
Japan	1988–1992	-6.3%	21.0%	I/B/E/S	Conroy and Harris (1995)
	1988–1992	-1.0%	14.6%	Toyo Kezai	"
	1989–1991	-3.7%	14.2%	I/B/E/S	Allen *et al* (1997)
	1989–1998	Not Reported	28.0%	I/B/E/S US Analysts	Higgins (2002)
	1992–1998	Not Reported	41.0%	I/B/E/S Japanese Analysts	Higgins (2002)
Korea	1989–1991	-4.7%	21.9%	I/B/E/S	Allen, *et al* (1997)
Hong Kong	1989–1991	-7.5%	16.1%	"	"
Malaysia	1989–1991	-3.7%	28.4%	"	"
Philippines	1989–1991	-14.2%	41.5%	"	"
Singapore	1989–1991	-8.7%	30.3%	"	"
Taiwan	1989–1991	-9.1%	25.5%	"	"
Thailand	1989–1991	-5.5%	35.3%	"	"

earnings. SURPE and SURPF are similar, expressing the forecast error as a percentage of the actual and forecast errors respectively. The absolute value of the denominator attempts to correct for negative earnings. They are signed, and thus provide information on the bias of the error. ABSPER is the absolute value of the forecast error scaled by the actual earnings. All four measures are reported on initially, as they have been used in the literature. Sebsequent chapters will focus on ABSPER and SURPE, the two most popular measures.

4.3.2 Descriptive statistics for the full sample

Descriptive statistics for the full sample for each of the error measures are presented in Table 4.2. Separate results for optimistic and pessimistic forecasts are presented in Table 4.2. Descriptive statistics for each of the error measures for the sample divided into industry groupings is presented in Table 4.3. For purposes of comparison international findings on forecast bias and the magnitude of the error are presented in Table 4.1

The signed error scaled by actual earnings, SURPE, is -11.2%. This would appear to be clear evidence of over optimism by analysts. The t-test is significant at the 5% level. SURPF is smaller at -1.3%, which confirms the evidence of over optimism (t-test not significant). Where the forecasts are on average larger than actual earnings, scaling by the forecast will result in smaller percentage errors. However, the median error in both cases is extremely small at - 0.2%. Optimistic forecasts occurred more frequently than pessimistic forecasts and had a larger median error value.

The results are not out of line with international findings either for the bias or the magnitude of the error. The bias is greater than has been found in Asian markets, while typically less than the bias found in the American markets. The comparison with the European markets depends on the study. Higgins (1998) found a forecast bias in each European country larger than the forecast bias found

Table 4.2: **Descriptive Statistics for the Full Sample** – The forecast in the month immediately prior to earnings announcement is compared to the realised earnings. Descriptive statistics are presented for the absolute error in cents per share (ABSURP), the percentage error scaled by forecast earnings (SURPF) and actual earnings (SURPE) and the absolute percentage error (ABSPER). The results presented are for the full sample, as well as the optimistic and pessimistic forecasts. The sample is for the years 1990–1999 inclusive.

	ABSURP	SURPF	SURPE	ABSPER
Full Sample	665			
Mean	18.5c	-1.3%	-11.2%	19.3%
Std. Dev	44.5c	44.9%	100.9%	99.6%
Median	4.5c	-0.2%	-0.2%	5.3%
Max	391.4c	767.4%	154.9%	2,226.9%
Min	0.0c	-345.3%	-2,226.9%	0%
t-statistic $H_0 : \mu = 0$		-0.75	-2.86	
Pessimistic Forecasts	276			
Mean		15.9%	9.7%	
Std. Dev		56.4%	16.5%	
Median		5.0%	4.7%	
Max		767.4%	154.9%	
Min		0.1%	01%	
Optimistic Forecasts	353			
Mean		-14.8%	-28.7%	
Std. Dev		29.9%	135.3%	
Median		-6.8%	-7.2%	
Max		0.1%	0.1%	
Min		-345.3%	-2,226.9%	

Table 4.3: **Descriptive Statistics for Individual Industries** – The forecast in the month immediately prior to earnings announcement is compared to the realised earnings. Descriptive statistics are presented for the absolute error in cents per share (ABSURP), the percentage error scaled by actual earnings (SURPE) and the forecast earnings (SURPF) and the absolute percentage error (ABSPER). The results for each of the industry groupings are presented separately.

	ABSURP	SURPF	SURPE	ABSPER
Industrials	487			
Mean	10.2c	-2.2%	-14.6%	21.1%
Std. Dev	18.7c	50.3%	116.5%	115.5%
Median	3.4c	-0.5%	-0.5%	4.4%
Max	183.9c	767.4%	142.2%	2,226.9%
Min	0.0c	-345.3%	-2,226.9%	0.0%
t-statistic $H_0 : \mu = 0$		-0.965	-2.767	
Financials	52			
Mean	19.3c	3.4%	-1.4%	11.7%
Std. Dev	31.4c	30.3%	19.1%	15.1%
Median	5.1c	0.6%	0.6%	4.6%
Max	163.4c	180.2%	64.3%	64.3%
Min	0.0c	-33.8%	-51.1%	0.0%
t-statistic $H_0 : \mu = 0$		0.809	0.529	
Resources	126			
Mean	45.0c	0.5%	-2.4%	15.3%
Std. Dev	66.7c	21.2%	30.3%	26.2%
Median	16.9c	0.0%	0.0%	8.5%
Max	391.4c	73.3%	154.9%	195.3%
Min	0.0	-96.7%	-195.3%	0.0%
t-statistic $H_0 : \mu = 0$		0.265	0.889	

the exception of 1992 the 15% bandwidth captured 70% of all forecasts. The distribution for Industrial and Financial stocks is similar, with the Industrial stocks having slightly more stocks in each bandwidth.

For management forecasts McDonald (1973) finds slightly more than 35% of the forecasts fall in a 5% error band, with 48.8% in a 10% band.

Richards (1976) found between 69% and 73% of all signed errors fell within the 10% bandwidth. Richards, Benjamin and Strawser (1977) found 32.6% in the 5% bandwidth, 50% in 10% and 61.2% in the 15%. Two percent of the errors fell outside the 100% bandwidth.

Dreman and Berry (1995) find that regardless of whether the error is scaled by earnings or forecast earnings, fewer than 30% of forecasts fall within a 5% bandwidth, with fewer than 45% within a 10% bandwidth.

4.3.4 Source of the errors

In the literature two methods of partitioning the mean squared forecast error are used to determine the source of the errors in earnings forecasts (naming conventions here follow Elton *et al* (1984)).

The mean squared forecast error is given by:

$$MSFE = \frac{1}{N}\sum_{i=1}^{N}(P_i - R_i)^2 . \tag{4.5}$$

P_i is the forecast change in earnings for firm i and R_i is the realised change in earnings for firm i.

Partition by level of aggregation

The first partition, by level of aggregation, divides the MSFE into the portion attributable to an inability to forecast at firm, industry and economy level. The

Table 4.5: **Partition by Level of Aggregation** – The values shown are the proportion of the Mean Squared Forecast Error attributable to an inability to forecast at the level of the full sample, the level of the industry subgrouping and at individual firm level.

	Full Sample	Industry	Firm	n
Full Sample	0%	0%	99%	665
1990	0%	0%	100%	43
1991	1%	8%	91%	50
1992	0%	1%	99%	51
1993	1%	0%	99%	51
1994	0%	5%	95%	54
1995	4%	0%	96%	56
1996	4%	6%	90%	66
1997	3%	3%	95%	78
1998	0%	0%	100%	98
1999	0%	7%	93%	113

partition by level of aggregation is given by:

$$MSFE = \left(\overline{P}-\overline{R}\right)^2+\frac{1}{N}\sum_{i=1}^{N}N_a\left[\left(\overline{P_a}-\overline{P}\right)-\left(\overline{R_a}-\overline{R}\right)\right]^2+\frac{1}{N}\sum_{i=1}^{N}\left[\left(\overline{P_i}-\overline{P_a}\right)-\left(\overline{R_i}-\overline{R_a}\right)\right]^2. \tag{4.6}$$

$\overline{P}$ and $\overline{R}$ are the mean forecast change and the mean actual change respectively for the full sample, $\overline{P_a}$ and $\overline{R_a}$ are the mean forecast and realised change in earnings for industry a and P_i and R_i are as previously defined.

The first term represents the portion of the squared forecast error attributable to analysts' inability to forecast growth for the full sample of forecasts under consideration. The second term, comparing industry means to the sample mean, is the inability to forecast at industry level. The third term is the portion of the error attributable to an inability to forecast at firm level. The percentage contribution of each of the three to the total error is given by dividing each by the MSFE.

Financial stocks are included from 1997 when there are seven firms in the industry category. This follows Elton *et al.* (1984). The full results are presented in Table 4.5

The largest part of the error is made at the firm level. The firm specific portion never accounted for less than 90% of the MSFE, and was 99% for the full sample. Errors made at the full sample level accounted less than 5% of the MSFE, with a mean of 0.3% for the full sample. The portion attributable to industry level errors is smaller for the full sample, but has a wider range of values for the individual range, from 0% to 8.1% of the MSFE.

These finding are consistent with Cragg and Malkiel (1968) who found that the industry level error was the smallest, followed by the full sample error. The largest part of the error for all forecasters, and never less than 50% of the error, was an inability to forecast at firm level. Overall they find that analysts do a poor job of forecasting long term growth rates. More recently Elton *et al* (1984) find that the inability to forecast at the level of the economy did not exceed 3% in the twelve months prior to the earnings announcement, and declined to 0.8% of the error in the month before the year end. The percentage attributable to the industry accounted for 15.5% in the last month of the firms' financial year, while the firm level accounted for 83.3% of the observed error.

The results for the JSE are consistent with those of other users of this decomposition in finding that the firm level error is the main contributor to the MSFE.

Partition by forecast characteristics

The second partition, by forecast characteristics, is based on the regression of the forecast earnings change on the actual earnings change. If analysts' forecasts are perfectly accurate, the fitted line will have an intercept of zero and a slope coefficient of one. Biased forecasts will have a non zero intercept.

$$MSFE = \left(\overline{P} - \overline{R}\right)^2 + (1-\beta)^2 \delta_P^2 + \left(1-\rho^2\right) \delta_R^2. \tag{4.7}$$

Where $\overline{P}$ and $\overline{R}$ are defined as before, β is the slope of the regression of the actual change in earnings on the forecast change in earnings, ρ is the correlation of forecast changes in earnings and actual changes in earnings and δ_P^2 and δ_R^2 are the standard deviation of the forecast change in earnings and the realised change in earnings respectively.

The first term of the output represents the extent to which the analysts errors were the result of bias. That is the extent to which the intercept differing from zero is responsible for the squared error. The second term measures the extent to which the squared error is a result of the regression slope differing from one. The last term is the portion of mean squared error attributable to random error. Again, the percentage contribution is calculated by dividing each component by the MSFE.

The results are presented in Table 4.6 for the full sample, as well as the sample divided by industry grouping and year.

For the full sample bias accounts less than 1% of the error, inefficiency for 3% and the random error term for 97%. In seven of the ten years, and for the Industrial and Resources firms the random term accounted for more than 90% of the error.

These results are similar to those of Elton *et al* (1984) who found that bias accounted for 0.9% of the total error, inefficiency for 3% and the random error term 96.1%. The findings of Crichfield *et al* (1978) are similar in finding that the random error is the most significant contributor to the total, 82%, but find higher values for bias, 13%, and inefficiency, 4.6%.

4.3.5 Regression analysis

The percentage change in earnings is regressed on the forecast percentage change. The results are presented in Table 4.7. The regression takes the following form:

Table 4.6: **Partition by Forecast Characteristics** – The values shown are the proportion of the Mean Squared Forecast Error attributable to bias (mean error different from zero), inefficiency (slope coefficient of regression of actual earnings change on forecast earnings change different from one) and random error.

	Bias	Inefficiency	Random Error	n
Full Sample	0%	3%	97%	665
Industrials	2%	4%	94%	487
Financials	1%	17%	82%	52
Resources	0%	2%	98%	126
1990	0%	43%	57%	43
1991	1%	1%	97%	50
1992	0%	39%	61%	51
1993	1%	9%	91%	51
1994	0%	0%	99%	54
1995	4%	19%	77%	58
1996	4%	2%	94%	69
1997	3%	7%	90%	78
1998	0%	1%	99%	98
1999	0%	1%	99%	113

Table 4.7: **Regression Analysis** – Results shown are for the regression of the actual percentage change in earnings on the forecast percentage change in earnings.

	Intercept	p Value H_0: α=0	Slope Coefficient	p Value H_0: β=1	R-Squared	Obs
Full Sample	-0.02	0.21	0.94	**0.03**	0.985	665
Industrials	-0.03	**0.04**	0.91	**0.00**	0.946	487
Financials	0.00	0.96	1.10	0.37	0.603	52
Resources	0.03	0.29	0.95	0.38	0.996	126
1990	-0.03	0.26	0.98	0.20	0.970	43
1991	-0.03	0.42	0.93	0.23	0.985	50
1992	-0.04	0.12	0.85	0.39	0.908	51
1993	-0.02	0.38	0.94	0.13	0.995	51
1994	0.00	0.92	0.78	0.13	0.966	54
1995	-0.03	0.53	0.86	**0.02**	0.885	58
1996	0.02	0.82	0.69	0.12	0.691	69
1997	-0.03	0.54	0.87	0.25	0.764	78
1998	0.04	0.18	0.99	0.18	0.861	98
1999	0.04	0.25	0.95	0.44	0.996	113

$$PercentageR_{it} = \alpha + \beta PercentageP_{it} + \epsilon_{it}. \qquad (4.8)$$

Where $PercentageR_{it}$ is the realized percentage change in earnings and $PercentageP_{it}$ the forecast percentage change in earnings for firm i at time t, α is a constant and ϵ_{it} is the error term.

The intercepts are mixed, positive in four years and negative in six. The only intercept significantly different from zero is for the Industrials, where the intercept is negative and significant at the 5% level, indicating a significant optimistic bias.

The slope coefficients are, with the exception of the Financial stocks, less than one. A forecast increase of 1% resulted in an actual increase of less than 1%.

These results are similar to those of Crichfield *et al* (1978) in finding mainly negative (but not significant) intercepts. Their slope coefficients were however consistently greater than one.

The results are consistent with Capstaff, Paudyal and Rees (2001) who found significantly negative intercepts, and slope coefficients less than one (except for Ireland where the intercept was positive). Their R^2 values are much lower, the greatest being in the Netherlands where the R^2 was 33%.

4.3.6 Forecast errors as the earnings announcement date approaches

Tables 4.8 and 4.9 show the change in forecast error during the forecast year for the absolute and signed percentage errors respectively. The results are for the full sample and for the industry groupings. The results for the individual years are presented in Appendix A.

ABSPER declines monotonically as the time to reporting draws nearer. While the data for months -12 to -10 may be tainted by old forecasts for the previous year this suggests that analysts correctly incorporate new information into their forecasts as the reporting year progresses. While the error does not decline as smoothly for each of the industries, the forecast at $t-1$ is never less accurate than the forecast in any of the preceding months.

In all years where the data exists, the forecast is always more accurate at $t-1$ than at $t-12$. There are only two years, 1992 and 1996, where the $t-9$ forecast (the first entirely untainted forecast) is on average less accurate than the $t-1$ forecast.

These results are consistent with Elton *et al* (1984) who regress their errors on time to maturity and find that forecasts become more accurate as the time to earnings announcement decreases. More recently Allen *et al* (1997) find similar results. They present SURPE and ABSPER at nine, six, three and one months before the reporting of earnings. The errors decline monotonically with the exception of the signed error of Taiwan between $t=-9$ and $t=-6$.

Table 4.8: **The Absolute Percentage Error the as Earnings Announcement Date Approaches** – The forecast in each month is compared to the realised earnings, and the absolute percentage error is calculated as $|(A_t - F_{t-n})/A_t|$ for $n = 1$ to 12 for each monthly observation.

t	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1
Full Sample	0.49	0.40	0.41	0.36	0.30	0.28	0.25	0.20	0.23	0.22	0.22	0.19
n	536	592	609	614	619	627	630	640	650	653	660	665
Industrials	0.47	0.43	0.43	0.39	0.33	0.31	0.27	0.21	0.24	0.25	0.24	0.21
n	391	434	444	447	452	457	460	469	479	481	484	487
Financials	0.18	0.16	0.15	0.15	0.13	0.14	0.15	0.13	0.13	0.12	0.12	0.12
n	43	47	50	50	50	52	52	52	52	52	52	52
Resources	0.69	0.36	0.44	0.30	0.27	0.24	0.22	0.21	0.24	0.17	0.15	0.15
n	102	111	115	117	117	118	118	119	119	120	124	126

Table 4.9: **The Signed Percentage Error the as Earnings Announcement Date Approaches** – The forecast in each month is compared to the realised earnings, and the absolute percentage error is calculated as $(A_t - F_{t-n})/A_t$ for $n = 1$ to 12 for each monthly observation.

t	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1
Full Sample	-0.24	-0.25	-0.26	-0.22	-0.17	-0.16	-0.14	-0.08	-0.05	-0.10	-0.09	-0.11
n	536	592	609	614	619	627	630	640	650	653	660	665
Industrials	-0.24	-0.31	-0.30	-0.28	-0.21	-0.19	-0.17	-0.09	-0.07	-0.12	-0.11	-0.15
n	391	434	444	447	452	457	460	469	479	481	484	487
Financials	0.13	0.08	0.04	0.04	0.02	0.00	0.01	-0.01	-0.02	-0.01	-0.01	-0.01
n	43	47	50	50	50	52	52	52	52	52	52	52
Resources	-0.43	-0.15	-0.22	-0.10	-0.11	-0.07	-0.10	-0.08	-0.03	-0.06	-0.04	-0.02
n	102	111	115	117	117	118	118	119	119	120	124	126

4.4 Conclusions

The South African evidence supports the recent international findings in the analysis of analysts' earnings forecasts. There is weak evidence of an optimistic bias in earnings forecasts. This bias is evident in both the mean signed error and the intercept of the regression of the actual earnings change on the forecast earnings change.

Forecasts for Resources and Financial stocks have a smaller mean error than Industrial stocks. Overall, the largest part of the forecast error is made at the level of the individual firm, and appears to be a random error.

While 49% of the observations fell within a 5% band around the realized earnings, for the full sample 25% of the observations fell outside the 15% bandwidth.

To the core question of whether the mean error is zero, the direct evidence comes from the signed percentage error and the regression analysis. The mean signed error scaled by actual earnings is significantly negative. The median error is however not distinguishable from zero. The intercept of the regression of the actual change on the forecast change is not significantly different from zero for the full sample. Overall these results suggest weak evidence of an optimistic bias.

forecasts and forecasts based on time-series extrapolations of past earnings. In the absence of a comprehensive database of management forecasts this study is limited to investigating time-series forecasts.

5.2 Prior research

A number of studies have compared the accuracy of analysts' forecasts to those based on extrapolations from past earnings. The evidence is mixed, with some finding analyst forecasts to be superior to forecasts based on mechanical time series forecasts (Crichfield *et al*, 1978; and Fried and Givoly, 1982). Others have found that analysts perform no better than simple time series extrapolations (Cragg and Malkiel, 1968; Elton and Gruber, 1972; and Bird, McElwee and McKinnon, 2000). By contrast others, like Deschamps and Mehta (1980), have compared different time-series forecasting techniques.

A related thread has considered the time-series properties of accounting income. Ball and Watts (1972) found that earnings were best described by a submartingale process - the best forecast for earnings at time t is the earnings at $t-1$. Brooks and Buckmaster (1976) extended this research by dividing their sample into strata. They found that while the submartingale process was the best description of the overall process, the outer strata, where earnings changes had been greatest in the previous periods, were best described by a mean reverting process. Albrecht, Lookabill and McKeown (1977) compare the forecasts of a random walk, a random walk with drift and Box-Jenkins. They find that for the full sample the random walk with drift provides a better fit than the random walk without drift. They found little difference in the predictive accuracy of the Box-Jenkins forecasts and the random walk forecasts.

The first paper of note that compares time series forecasts to analyst forecasts comes from Elton and Gruber (1972). They compare the accuracy of analysts' forecasts to nine mechanical forecasting techniques based on exponentially weighted moving averages, naïve models, simple moving averages and regressions. Their analysis covers the period 1964–1966, with forecasts taken from three financial institutions. Their methodology is based on Theil's U statistic. They calculate a U statistic for each of the earnings realizations, and then compare the U statistics of the different forecasting methods. While they conclude that analysts perform no better than time series models, it should be noted that they compare the individual analysts to the forecast technique found ex post to be the most successful at forecasting earnings.

Crichfield *et al* (1978) improve on the methodology of Elton and Gruber (1972). Their analysis is based on forecasts for 46 companies taken from the Standard and Poor's Earnings Forecaster for the period 1967–1976. Forecasts were averaged where more than one appeared in a given month. Their sample selection method has two problems. First, they select a few consecutive pages from two starting points, and then eliminate all firms for which there is no forecast in one or more of the ten years of the study. Aside from grouping that may occur with firms appearing on adjacent pages on a non random basis, eliminating firms with missing observations may lead to survivorship bias. Modifying Theil's U statistic to allow for a direct comparison between analysts and the time series forecasts they compare the analysts' forecasts to five models; a naïve no-change model, a three year moving average and three models based on quarterly earnings. For each of the two models based on full year earnings the analysts' forecasts where more accurate than the time series model in each of the ten years of the study.

Brown and Rozeff (1978) use Box-Jenkins forecasts as their primary basis of evaluating Value Line forecasts. In testing they find the Box-Jenkins forecasts to be superior to those produced by, 'ad hoc naïve time series models'. For a random selection of 50 shares they are able to reject the null hypothesis of the median Value Line forecast being equal to the time series forecasts, and find the Value Line forecasts to produce smaller errors. In the individual years of the study forecasting at a one month horizon the Value Line forecasts are significantly better than the Box Jenkins at a 1% level in two years, at a 5% level in one year and are indistinguishable in the fourth. Their testing is based on the Wilcoxon Signed Rank test. They criticize the use of Theil's U in Elton and Gruber (1972) on the basis that significance testing is not possible, and that direct testing is preferable to ranking and comparing U statistics.

Fried and Givoly (1982) analyse forecasts from the Standard and Poor's Earnings Forecaster, including all firms for which forecasts were available for the eleven year period from 1967–1979. They compute the consensus forecast as the mean of individual forecasts published. Two time series forecasting methods are used. The first, following the results of Albrecht *et al* (1977), forecasts future earnings as the current earnings plus the six year arithmetic average of earnings growth (martingale with trend). Following Brooks and Buckmaster (1976) the sample was stratified and an exponential smoothing model was used for firms in the extreme strata. They find that the analysts' mean absolute percentage errors are smaller than either of the time series models, however the analysts produced more biased forecasts than either of the time series models. In net their evidence is mixed.

Bird, McElwee and McKinnon (2000) examine earnings forecasts for firms included in the Morgan Stanley Capital Index World universe as at June 1998. The forecasts are one and two year forecasts taken from the I/B/E/S database for the period January 1987–June 1998. They compare the analysts' forecasts to a naïve no-change forecast, and a forecast based on extrapolating from the previous five years of earnings. With the exception of Norway they find that the naïve no-change model performs at least as well as the analysts.

On the comparison of different time-series forecasting techniques, with no reference to analysts, Deschamps and Mehta (1980) find the submartingale to be dominated by the constant growth, Box-Jenkins and a third method that combines the submartingale and constant growth models. They were unable to separate between the last three models.

5.3 Methodology

The first part of this analysis follows Crichfield *et al* (1978) in using a modified form of Theil's U statistic to compare the accuracy of analysts' forecasts relative to those produced by time series forecasts.

The standard Theil's U statistic is :

$$U = \frac{\sum_{i=1}^{N} (P_i - R_i)^2}{\sum_{i=1}^{N} R_i^2}. \tag{5.1}$$

The forecast change, P_i, is the forecast earnings less the earnings in the previous period and the realized change or $(F_{it} - A_{it-1})$. The actual change R_i is the actual earnings in the current period less the earnings in the previous or $(A_{it} - A_{it-1})$. Substituting this into 5.1 gives:

$$U = \frac{\sum_{i=1}^{N} \left((F_{it} - A_{i,t-1})^2 - (A_{it} - A_{i,t-1}) \right)}{\left(\sum_{i=1}^{N} A_{it} - A_{it} \right)^2}. \tag{5.2}$$

This can then be simplified to:

$$U = \frac{\sum_{i=1}^{N} (F_i - A_i)^2}{\sum_{i=1}^{N} (A_i - A_{it-1})^2}. \tag{5.3}$$

Theil's U thus solves to being the ratio of the squared actual change in earnings to the squared forecast change in earnings, and in this form compares the analysts' error to the error that the analysts would have made had they forecast a repeat of last year's earnings. By replacing the A_{it-1} with the forecast of a time series forecast a direct comparison can be made between the analysts' error and the error that would be made by making the time series generated forecast. This is the method of testing used by Crichfield *et al* (1978). There is still no statistical test to determine the significance of the results, but this methodology does offer a direct comparison of the accuracy of the analysts' forecast and the time series forecast.

A value of zero indicates that the analyst's forecast is the same as the realized earnings. A value of one indicates that the two squared errors are equal, and the error made by the analysts is equal in magnitude to the error made by the time-series forecast. A value between zero and one indicates an analyst forecast error of lesser magnitude than the time-series forecast error, and a value greater than one indicates an analyst error greater in magnitude than time-series forecast. Theil's U only considers the magnitude and not the direction of the relative errors.

The forecasting techniques used are the naïve no-change model, a three and five year mean earnings, the same growth (percentage and in Rands) as the last change and a three and five year mean percentage change (arithmetic and geometric).

The models are:

1. M1: naïve, no-change model.

$$F_{M1t} = A_{t-1}. \tag{5.4}$$

2. M2: Same percentage growth as previous period.

$$F_{M2t} = A_{t-1}\left[1 + \frac{A_{t-1} - A_{t-2}}{A_{t-2}}\right]. \tag{5.5}$$

3. M3: Same growth in Rands as previous period.

$$F_{M3t} = A_{t-1} + (A_{t-1} - A_{t-2}). \tag{5.6}$$

4. M4: Average of the last three years actual earnings.

$$F_{M4t} = \frac{A_{t-1} + A_{t-2} + A_{t-3}}{3}. \tag{5.7}$$

5. M5: Average of the last five years actual earnings.

$$F_{M5t} = \frac{A_{t-1} + A_{t-2} + A_{t-3} + A_{t-4} + A_{t-5}}{5}. \tag{5.8}$$

6. M6: Same growth as three year mean arithmetic growth.

$$F_{M6t} = A_{t-1}\left[1 + \frac{1}{3}\sum_{n=1}^{3}\left(\frac{A_{t-n} - A_{t-n-1}}{A_{t-n-1}}\right)\right]. \tag{5.9}$$

7. M7: Same growth as three year mean geometric growth.

$$F_{M7t} = A_{t-1}\left[1 + (A_{t-1} - A_{t-3})^{\frac{1}{3}}\right]. \tag{5.10}$$

8. M8: Same growth as five year mean arithmetic growth.

$$F_{M8t} = A_{t-1}\left[1 + \frac{1}{5}\sum_{n=1}^{5}\left(\frac{A_{t-n} - A_{t-n-1}}{A_{t-n-1}}\right)\right]. \tag{5.11}$$

9. M9: Same growth as five year mean geometric growth.

$$F_{M9t} = A_{t-1}\left[1 + (A_{t-1} - A_{t-5})^{\frac{1}{5}}\right]. \tag{5.12}$$

Where F_{MNt} is the forecast produced by model MN for the full year earnings announced at time t. A_{t-n} is the full year earnings for the year $t - n$ where t is measured in years.

The analysts' forecast from the month before earnings are announced are compared to the time series forecast. The results for the full twelve months prior to earnings announcement are presented for the naïve, no-change forecast only.

Finally the accuracy of the analysts' forecasts, measured by the absolute percentage error and the absolute error in cents per share, are compared to the time series forecasts using the Wilcoxon Signed Ranks test (following Brown and Rozeff (1978), see also Conover (1980) for a fuller discussion). The Wilcoxon Signed

Ranks test is a non-parametric test that compares medians for equality. It has the advantage over the paired t-test that it is robust in the face of outliers, makes no assumptions about the distribution of the errors and is robust to the error definition used, Brown and Rozeff (1978). The test cannot be performed with the signed percentage error - a negative difference between the analyst and time series forecasts is ambiguous when the analyst errors and time series errors can have different signs.

The one-sided null hypothesis is that the analysts' median error is equal to or greater than the median error of the time series forecasting technique.

H_0: E (Analysts' Error) $\geq$ E (Time series forecast error)

H_1: E (Analysts' Error) $<$ E (Time series forecast error)

The test statistic is calculated as:

$$T = \frac{\sum_{i=1}^{n} R_i}{\sqrt{\sum_{i=1}^{n} R_i^2}}. \tag{5.13}$$

Where R_i is the signed rank of the absolute differences between the analyst forecast and the time-series forecast. Given the size of the sample and the fact that there are ranking ties the significance of T is taken from the standard normal distribution (Conover,1980:282).

The analysts are only compared to a relatively short list of possible time series forecasting techniques. Box-Jenkins would have been preferred however the time series of earnings forecasts required to produce earnings forecasts is not available. By way of comparison Brown and Rozeff (1978) use a twenty year earnings history in estimating Box Jenkins forecasts.

5.4 Results

The results of the testing for the full sample and for each of the industry groups are presented in Table 5.1. Results for the full sample for each of the years are presented in Table 5.2.

For the full sample analysts clearly dominate the time series forecasts. The best of the time series forecasters is the five-year mean geometric growth model. The U statistic of 0.97 indicates that the analysts are at best fractionally better than, and more likely indistinguishable from, the time-series forecasts. The relative ability of this forecasting technique appears to come from its ability to forecast earnings for Financial stocks only. The U value is in excess of 0.99 for the Financials and less than 0.1 for the Resources and Industrial stocks. The forecasts based on the five year arithmetic growth produce similar results. The overall U is lower, but still higher than any of the other techniques, and the U value is again a result if its ability to forecast earnings for Financial stocks as well as the analysts. Further investigation of the five year growth forecasting techniques indicates that not only is the superiority of these techniques based on an ability to forecast earnings for Financial stocks, but that this ability is only evident in 1999. The best overall time series forecasting method is using the naïve no-change forecast. It provides the highest U values for both Industrial and Resource stocks, and fourth highest for the Financial stocks. Ranking the techniques for each of industrial groupings and summing the rankings provides some measure of the ability of a time series forecasting technique to forecast earnings. The lowest three summed ranks are, in order, No-Change, Same Rand Growth and the Three Year moving average. For the full sample the no-change forecast is ranked third, beaten by the two five year mean growth models (by virtue of their ability to forecast earnings for Financials).

Intra-year comparison of analyst and time-series forecasts

The second part of this analysis looks at the accuracy of analysts forecasts within the financial year. The comparison is limited to the naïve no-change forecast. The results are presented in Table 5.3. Theil's U statistics are presented for each of the

Table 5.1: **A Modified Thiel's U Comparison of Analyst and Time-Series Forecasts** – The table contains the results for the comparison of the analysts' forecasts and the time-series forecasts. The test statistic, based on a modified form of Theil's U statistic, is calculated as U = squared analysts' error/squared forecasting technique's error. Results are presented for each of the techniques for each of ten years of the study. The sample size in any period is n.

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
No-Change	**0.44**	**0.10**	**0.99**	**0.18**	**0.14**	**0.34**	**0.08**	**0.21**	**0.22**	**0.14**
n	43	50	51	51	54	58	69	78	98	113
%g	**0.21**	**0.01**	**0.39**	**0.24**	**0.08**	**0.18**	**0.03**	**0.09**	**0.06**	**0.00**
n	40	47	51	51	51	57	65	74	86	105
Rand g	**0.56**	**0.02**	**0.25**	**0.23**	**0.07**	**0.14**	**0.09**	**0.22**	**0.07**	**0.23**
n	40	47	51	51	51	57	65	74	86	107
3 Year MA	**0.16**	**0.27**	**0.28**	**0.07**	**0.09**	**0.36**	**0.05**	**0.08**	**0.15**	**0.09**
n	37	43	48	51	51	57	63	67	81	88
5 Year MA	**0.11**	**0.08**	**0.32**	**0.11**	**0.06**	**0.30**	**0.05**	**0.06**	**0.08**	**0.08**
n	33	36	43	44	48	55	57	58	70	75
3 Year Growth (Ari)	**0.12**	**0.01**	**0.10**	**0.11**	**0.12**	**0.26**	**0.07**	**0.26**	**0.07**	**0.01**
n	33	39	44	48	51	55	57	64	74	83
3 Year Growth (Geo)	**0.03**	**0.01**	**0.10**	**0.13**	**0.11**	**0.30**	**0.07**	**0.29**	**0.05**	**0.13**
n	33	39	44	48	51	55	57	64	75	83
5 Year Growth (Ari)	**0.26**	**0.01**	**0.06**	**0.06**	**0.12**	**0.32**	**0.07**	**0.27**	**0.11**	**0.94**
n	35	36	37	40	44	50	55	57	66	73
5 Year Growth (Geo)	**0.61**	**0.02**	**0.02**	**0.03**	**0.07**	**0.46**	**0.05**	**0.11**	**0.17**	**0.99**
n	36	37	37	40	44	50	55	57	66	73

Table 5.2: **A Modified Thiel's U Comparison of Analyst and Time-Series Forecasts Across Industry Groups** – The table contains the results for the comparison of the analysts' forecasts and the time-series forecasts. The test statistic, based on a modified form of Theil's U statistic, is calculated as U = squared analysts' error/squared forecasting technique's error. Results are presented for each of the techniques for each of the industry groupings in the sample. The sample size is n.

	Full Sample	Industrials	Financials	Resources
No-Change	**0.19**	**0.26**	**0.15**	**0.18**
n	665	487	52	126
Same % Growth	**0.00**	**0.18**	**0.00**	**0.08**
n	627	457	47	123
Same Rand Growth	**0.14**	**0.24**	**0.07**	**0.14**
n	629	459	47	123
3 Year MA	**0.11**	**0.13**	**0.07**	**0.11**
n	586	425	40	121
5 Year MA	**0.09**	**0.09**	**0.04**	**0.09**
n	519	369	35	115
3 Year Growth (Ari)	**0.03**	**0.09**	**0.00**	**0.09**
n	548	392	38	118
3 Year Growth (Geo)	**0.06**	**0.05**	**0.23**	**0.06**
n	549	393	38	118
5 Year Growth (Ari)	**0.78**	**0.08**	**0.99**	**0.11**
n	493	349	31	113
5 Year Growth (Geo)	**0.97**	**0.07**	**1.00**	**0.08**
n	495	351	31	113

Table 5.3: **Comparison of the Analysts' and the No-Change Forecasts' Errors for the Twelve months prior to Earnings Announcement** – Analysts' forecasts are compared to forecasts of no-change in earnings using Theil's U statistic. A value of zero indicates that the analysts have forecast the earnings perfectly accurately. A value of one indicates that the magnitude of the analysts' error is identical to the magnitude of the error made by a no-change forecast. A value of between zero and one indicates that the magnitude of the analysts' error is less than that of the no-change forecast while a value greater than one indicates that the analysts' error was of greater magnitude than the time-series forecast. The results are presented for the full sample and for each of the industry groups. The sample size is n.

t	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1
Industrials	**1.23**	**1.10**	**0.99**	**0.94**	**0.86**	**0.88**	**0.71**	**0.70**	**0.69**	**0.34**	**0.33**	**0.26**
n	391	435	444	447	452	457	460	469	479	481	484	487
Resources	**0.82**	**0.53**	**0.48**	**0.46**	**0.53**	**0.39**	**0.29**	**0.23**	**0.22**	**0.29**	**0.16**	**0.18**
n	102	111	115	117	117	118	118	119	119	120	124	126
Financials	**0.22**	**0.16**	**0.13**	**0.17**	**0.13**	**0.19**	**0.17**	**0.19**	**0.19**	**0.18**	**0.14**	**0.15**
n	43	47	50	50	50	52	52	52	52	52	52	52
Full Sample	**0.83**	**0.59**	**0.53**	**0.51**	**0.55**	**0.45**	**0.35**	**0.30**	**0.29**	**0.29**	**0.18**	**0.19**
n	536	593	609	614	619	627	630	640	650	653	660	665

twelve months before earnings are announced. The evidence for the full sample is clear, analysts' twelve month forecasts are more accurate than the no-change forecast and the analysts' forecasts improve as the earnings announcement date approaches. This superiority is seen in the Resources and Financial stocks, where the evidence is largely the same as for the full sample. This superiority is not seen in the Industrial stocks, where the naïve forecasts' error is less than that of the analysts' for a twelve and eleven month horizon. From this point the analysts' forecasts improve and are by the end of the year superior to those of the naïve no-change model. It is clear from the results that analysts improve their forecasts as the earnings announcement date approaches. If this observation is explained by analysts modifying their forecasts as new information arrives in the market, then one would expect analysts' superiority over time-series forecasts to increase during the firm's financial year.

Results of the Wilcoxon Signed Ranks Test

The results of the Wilcoxon Signed Ranks tests are presented in Table 5.4. The results indicate clear analyst superiority over the time series forecasts considered. The t-statistics are all greater than ten and the null hypothesis that the median analyst forecast error is less than or equal to the median time series forecast error can be rejected for all time series forecasting techniques at the 1% significance level.

Descriptive Statistics

Brief descriptive statistics are presented for each of the time-series forecasting techniques with the corresponding results from the analysts in Table 5.5. Analysts produce smaller mean errors and standard deviations of errors in cents per share and in absolute percentage error than the time-series forecasts. Analysts' forecasts do not have the minimum bias of the techniques examined. The

Table 5.4: **Wilcoxon Signed Ranks Test Results** – The results presented are for a one-tailed test with the null hypothesis that the analysts' errors are greater than or equal to those of the time-series forecasting techniques. The alternative hypothesis is that the analysts' errors are smaller than those made by the time series forecasting techniques. The null hypothesis is rejected at the 0.1% significance level for T values greater than 3.090. The t-statistics are presented for each of the forecasting techniques with p values in parentheses. The sample size in each case is n.

	Abs (CPS)	n	ABSPER	n
No-Change	**15.59**	664	**16.21**	664
	0.00		0.00	
3 Yr MA	**16.88**	585	**18.53**	585
	0.00		0.00	
Same % Gro	**14.39**	630	**14.58**	629
	0.00		0.00	
Same Rand Gro	**13.75**	628	**13.76**	627
	0.00		0.00	
3 Yr Ari	**14.49**	549	**14.76**	549
	0.00		0.00	
3 Yr Geo	**15.55**	549	**15.88**	549
	0.00		0.00	
5 Yr MA	**16.96**	518	**17.33**	519
	0.00		0.00	
5 Yr Ari	**14.07**	486	**14.12**	486
	0.00		0.00	
5 Yr Geo	**14.03**	488	**14.17**	488
	0.00		0.00	

Table 5.5: **Descriptive Statistics for Time-Series Forecasting Techniques**

		Mean	Std Dev	Median	n
ABS_ERROR (c)	Analysts' Error	17.5	36.8	4.49	665
	No-Change	45.0	81.4	17.9	665
	3 Yr MA	65.1	112.5	26.3	585
	Same % g	82.5	640.8	16.7	630
	Same Rand g	50.3	99.9	13.9	629
	3 Yr Ari	75.1	260.9	19.0	549
	3 Yr Geo	76.6	170.6	22.2	549
	5yr MA	81.8	131.1	37.7	518
	5 Yr Ari	69.6	126.6	18.8	486
	5 YR Geo	76.1	143.0	21.9	488
ABSPER (%)	Analysts' Error	19	100	5	665
	No-Change	53	189	21	665
	3 Yr MA	62	199	31	585
	Same % g	92	588	18	630
	Same Rand g	71	328	14	629
	3 Yr Ari	70	366	17	549
	3 Yr Geo	73	463	22	549
	5yr MA	63	172	38	518
	5 Yr Ari	57	297	17	486
	5 YR Geo	60	292	22	488
PERA (%)	Analysts' Error	-11	101	0	665
	No-Change	0	196	15	665
	3 Yr MA	-12	106	0	585
	Same % g	-26	594	-4	630
	Same Rand g	-8	335	1	629
	3 Yr Ari	-28	372	-4	549
	3 Yr Geo	-42	467	-8	549
	5yr MA	13	183	34	518
	5 Yr Ari	-27	302	-4	486
	5 YR Geo	-25	297	7	488

no-change forecast has a mean bias of zero while the forecast of the same Rand growth as the previous period, while optimistically biased, is less biased than the analysts.

It is clear from these results that analysts' forecasts are more accurate than those produced by naïve extrapolative techniques.

5.5 Forecasting turning points

None of the time-series forecasting models can forecast a change in the direction of earnings from the last period. Perhaps in this area analysts enjoy the greatest advantage over time-series models. As prices are particularly sensitive, in an earnings or dividend discount model, to changes in the direction of earnings growth, correctly forecasting even a modest percentage of all turning points may justify the expense of employing analysts.

In the full sample there were 178 occasions where the direction of the growth in earnings was different to the previous period. On 95 occasions positive earnings growth became negative earnings growth. On 83 occasions negative earnings growth became positive earnings growth. In total analysts correctly forecast the direction change correctly in 141 occasions. In doing so they missed four upturns and 33 downturns.

5.6 Conclusions

The analysts produced forecasts clearly superior to the time-series models tested. This is confirmed with the Thiel's U statistics and the results of the Wilcoxon Signed Ranks test. The advantage increases as the time to earnings announcement date approaches.

Chapter 6

Forecast Predictability – A Balanced Panel Approach [1]

6.1 Introduction

This part of the analysis is concerned with the predictability of analysts' forecast errors, and specifically whether there are patterns in the errors. Given that the results of the decomposition of the mean squared forecast error into economy, industry and firm components showed that the largest part of the error is made at the firm specific level, the focus is on the ability of firm specific characteristics to predict forecast errors. Two macroeconomic variables are included, with the expectation that they will not yield significant results.

6.2 Prior research

The earliest suggestion of a multifactor model to explain analysts' forecast errors is McDonald (1973:510), 'research is needed to determine what factors affect

[1] The results contained in this chapter were presented at the Southern African Finance Association annual conference in January 2004.

a firm's ability to predict earnings. These factors include both those factors endogenous and exogenous to the firm.' No specific factors are proposed.

The earliest suggestion of specific factors is Albrecht, Johnson, Lookabill and Watson (1977). They propose, without empirical testing, a model where the forecast accuracy is a function of earnings variability, firm age, firm size, the industry to which the firm belongs, the time to earnings announcement, the calendar year of the forecast and the individual forecaster. A final factor is designated as 'detail of information', and is intended to capture the quality of information available about the firm.

Foster (1986) follows Albrecht *et al* (1977) in proposing, without any supporting empirical evidence, factors that may impact on the accuracy of an analyst's forecast. Foster's proposed factors are similar to those of Albrecht *et al* (1977); time to the earnings announcement, the industry, the size of the firm, the number of lines of business, the variability of the firms' past earnings, the year of the forecast and a dummy variable to reflect accounting choices and the individual forecaster.

The first empirical testing in this area is reported by Brown, Richardson and Schwager (1987). Following the evidence that analysts forecast earnings more accurately than time-series forecasting models, they observe that, 'The source of this apparent superiority, however, is not well understood'(p49). Rather than look at the predictability of analysts' forecast errors directly, the authors look at the factors as explanations for the superiority of analysts' forecasts over time-series model forecasts. Their dependent variable is the ratio of the variance of the error made by no-change forecast to the variance of the analysts' forecast error. Two sources of forecasts were tested, first the Value Line forecasts, as proxies for single analyst forecasts (the Value Line forecasts are made by up to two analysts) and the I/B/E/S consensus forecasts. The explanatory variables are firm size (equity plus long term debt), the dispersion of analysts' forecasts and the number of lines of business. The expectation is that analyst superiority is positively related to firm size, negatively related to the dispersion of forecasts and positively related to the number of lines of business. Quarterly dummies are used to control for the time to earnings announcement, given the well established finding that the accuracy of analysts' forecasts improve as the announcement of full year earnings approaches. A statistically significant relationship is found between analyst superiority and firm size and analyst forecast dispersion. Both slope coefficients have the expected sign. The relationships are estimated by pooling the data for a sample period from 1977–1982.

Kross, Ro and Schroeder (1990) follow Brown, Hagerman and Zmijewski (1987) in estimating a model to explain the superiority of analysts forecasts over time-series model forecasts. They add two previously unexplored explanatory variables, the number of column inches in the Wall Street Journal and the historical error based on time series forecasts. The column inches in the Wall Street Journal are used as a proxy for the amount of information available about the firm. In addition their explanatory variables include firm size (equity only), the number of lines of business, the time to earnings announcement and a dummy variable for the industry. Their dependent variable is the difference between the analysts' absolute error in cents per share less the absolute value of the error in cents per share made by a time series model based on Brown and Rozeff (1978). The problem with an unscaled error definition is that if a stock in their sample split its stock ten for one, the split would change the dependent variable without changing firm size or the number of column inches in the Wall Street Journal Index.

The authors find that the analysts' advantage over time series forecasts to be positively and significantly related to the variance of previous errors, Wall Street Journal exposure and the time to earnings announcement, with a weaker correlation with lines of business. No significant relationship is found between analysts' advantage and the size of the firm. While their results contradict those of Brown *et al* (1987) the unscaled dependent variable renders their results unreliable.

Das *et al* (1998) investigate the predictability of analysts' forecast bias, measured as the signed error scaled by the share price. They estimate a model where the bias is a function of the unpredictability of earnings, the analyst following, firm size, the Value Line timeliness rating of the forecast and two interaction variables, one for the interaction between the analyst following and the unpredictability of earnings and the other the interaction between firm size and the unpredictability of earnings. The timeliness variable is the Value Line recommendation for the stock, with values ranging between one for a strong buy and five for a strong sell. Their sample covers 241 firms for the period 1989–1993. The unpredictability proxy is based on the errors made by time series forecasts, where the forecasting method is chosen based on its fit with the firm's historical earnings realizations.

Analyst following and firm size are found to be highly (positively) correlated while neither is found to be correlated with the unpredictability measure. A significant negative relationship is found between earnings predictability and analyst bias. A significant positive relationship is found between firm size and analyst bias for the one and two quarter horizons. The interaction variable between following and unpredictability is significant for the one and two quarter horizons and finally the timeliness is significantly negatively related for the two, three and four quarter horizon. The timeliness result indicates a negative relationship between bias and the Value Line recommendation.

Using a sample of 2,078 firm years from Fortune 500 companies from the period 1976 to 1983 Parkesh, Dhaliwal and Salakta (1995) use a latent variables approach to explain forecast errors as a function of production, financing and investing uncertainty and the availability of information about the firm. They find a significant positive relationship between variables proxying for firm risk (variables include debt-to-equity ratio, the concentration of ownership and two variables based on the variability of accounting and market returns respectively), and a negative relationship between the error term and the availability of information.

Expectations from the asset pricing literature

In addition to the factors used in the above studies, empirical research into stock returns suggests the use of the ratio of the book value of assets to the market price of the stock as an explanatory variable. Two possible alternatives to the book-to-market ratio are the price-to-earnings ratio and the dividend yield. This is based on the empirical research of Fama and French (1992) who find that firm size and the book-to-market ratio predict stock returns. The price-to-earnings ratio and dividend yield capture much of the same variation of returns as the book-to-market ratio, but are found to be subsumed by the book-to-market ratio in asset pricing tests. Typically referred to as the size and value effects, there are positive abnormal returns to small firms over large firms and positive abnormal returns to value stocks (high book-to-market) where normal returns are based on the CAPM.

This finding has been confirmed over different periods in the U.S. as well as in the rest of the world. Davis, Fama and French's (2000) study covers the longest time period in the U.S., with results significant over a 68 year period. Fama and French (1998) find similar results for the value premium (high book-to-market ratio, high earnings yield and high dividend yield) in a sample of twelve countries' stock returns. These included European and Asian countries as well as Australia.

The current debate in the asset pricing literature is over the source of these findings. The discussion that follows is summarized from Davis, Fama and French (2000). The first school interprets these results as peculiar to the time period in question, simply a result of data mining that will not be repeated in the future. Proponents of this line include Black (1993) and MacKinlay (1995). This line of argument is partly answered by evidence from outside the U.S., particularly Fama and French (1998). Chan, Hamao and Lakonishok (1991) find a positive relationship between stock returns and the book-to-market ratio in the Japanese market for the period 1971 to 1988 and a negative relationship between returns and firm size. Earnings yield is positively related to returns when tested alone or with firm size, but is entirely subsumed by the book-to-market ratio when included.

The second school, Fama and French (1992) and Davis, Fama and French (2000) among others, find the effects to be a reward for bearing risk that is not captured by the CAPM, and that a fully specified risk model for a portfolio needs to take into account these factors.

The third school is the behaviorists. DeBondt and Thaler (1987), Haugen (1995) and Lakonishok, Shleifer and Vishny (1994) see the abnormal returns as a result of investor overreaction and irrationality. Analysts extrapolate good (and poor) performance too far. When this overreaction is corrected weak firms (typically with high book-to-market ratios) outperform while low book-to-market firms underperform.

Finally, Daniel and Titman (1997) present evidence that the results are a result of the firm's characteristics rather than the estimated coefficient. The previous work relied on a firm's covariance with portfolios that attempted to capture the premium attributable to the effect under study. For the size effect firm covariances were measured against a portfolio whose return was equivalent to a long position in a portfolio of small capitalization shares and a short position in a portfolio of large capitalization shares. Similar portfolios were constructed for the other effects. Daniel and Titman (1997) argue that the source of the value and size premiums are not the covariance of firm returns with the benchmark portfolio returns, but rather the actual size and book-to-market ratio of the firm in question.

It is unlikely that the data sample in this study is large enough, in either the cross section or the time series, to discriminate between the competing hypotheses. It should, however, be possible to separate the efficient markets from inefficient markets arguments with a study that analyses the errors that forecasters make in forecasting firm earnings. If firm size and the price-to-earnings ratio are priced risk factors, then they will be uncorrelated with forecast error, as the analysts, with a complete understanding of the earnings generation process will have accounted for these factors in estimating earnings.

If, on the other hand, a relationship is found between the price-to-earnings ratio of the firm and the size of the error, then this would suggest that analysts make systematic errors in forecasting earnings for growth stocks relative to value stocks. The value effect will then be captured in bias regressions where optimism is expected in forecasting for stocks with high price-to-earnings ratios and pessimism in forecasting for stocks with low price-to-earnings ratios.

The argument as to the size premium is more difficult. The size factor will be captured, if at all, by the absolute percentage errors – here the interest is simply whether large stocks are associated with smaller mean errors (and consequently have a lower expected return).

6.3 Data

The purpose is to test the relationship between the forecast error and the forecast bias and candidate explanatory variables. The full set of variables is discussed below, followed by a description of the data used in the testing.

6.3.1 Dependent variables

Two dependent variables are considered, the absolute percentage error and the signed error scaled by actual earnings.

$$ABSPER_{it} = \left| \frac{A_{it} - F_{i,t-1}}{A_{it}} \right|. \tag{6.1}$$

$$PERA_{it} = \frac{A_{it} - F_{i,t-1}}{A_{it}}. \quad (6.2)$$

ABSPER is the absolute percentage error as used previously. PERA differs from the SURPE used previously in that the absolute value of the denominator is not used.

PERA is similar to the dependent variable used by Das, *et al* (1998). Their measure of bias differs from PERA in that it is scaled by share price rather than actual earnings.

6.3.2 Independent variables

The explanatory variables are given below. The only prior expectation is that the macroeconomic variables will not provide significant explanations of the errors. This is based on the finding that the largest portion of the mean squared forecast error is attributable to an inability to forecast at the firm level. A full list of explanatory variables is presented in Table 6.1.

Beta (BETA)

If beta is an accurate measure of systematic risk, and earnings errors are the realized expression of risk, then the expectation is that high beta firms should on average be associated with high forecast errors. The coefficient for BETA would then be expected to be significant for the absolute percentage error, but not for the signed percentage error, as risk simply implies a large mean error, rather than predicting the direction of the error.

Firm Size (LNMC)

The size factor is included for two reasons, first that it is a proxy for the importance of the firm to the investment community and second, that it has been found to be associated with abnormal returns as discussed above. The relative importance of a firm to the investing community would then appear to be a reasonable proxy for the quality of the information available about a firm. This quality of information will typically reflect greater analyst coverage, more media attention. The firms in the sample vary in size between R52.5 million and R68 billion, with a median market capitalization of R3 billion. If analyst coverage is important for forecast accuracy and size is a reasonable proxy for analyst coverage then one would expect a negative relationship between firm size and absolute percentage forecast error. It is not expected that there is a relationship between the direction of the error and the size of the firm.

Earnings Yield (EY)

In the absence of an available measure of the book value of assets to the market value of the firm, earnings yield and dividend yield are used as proxies for value, being a measure of the cheapness of the share relative to current earnings. This variable is included because of its demonstrated relationship to stock returns. The expectation is that if the earnings yield predicts returns because analysts overestimate earnings for firms with high earnings yields and underestimate earnings firms with low earnings yields, then the relationship to error is likely to be captured in the bias regression, where a positive slope is expected.

Dividend Yield (DY)

If the book-to-market premium is linked to earnings, and specifically the difficulty of forecasting, then this the Dividend Yield is expected to capture the same variation as the Earnings Yield. A positive relationship between Dividend Yield and the bias of analysts' errors is expected.

Historical Growth (HGRO)

The historical growth factor is an attempt to capture any effect that earnings growth has on forecast accuracy.

Inflation (INF)

For reasons discussed previously in the decomposition of the mean squared forecast error the rate of inflation is not expected to have significant forecasting power in predicting forecast errors.

Interest Rate (PRIME)

Similar to the rate of inflation, the prime interest rate is not expected to have any significant forecasting power.

Lagged Forecast Error (ABSPER_L, PERA_L)

This is a direct measure of how difficult it is to forecast earnings for a firm. In the literature the difficulty of predicting earnings has either been included in the dependent variable where the analysts' errors (or variability) are compared to the errors made by a time-series forecasting model, as in Brown *et al* (1987) or Kross *et al* (1990) or through the inclusion of an independent variable based on the forecasting errors that would have been made by time-series models in forecasting previous earnings for the firm in question, Das *et al* (1998). Using the analysts' lagged error as a measure of difficulty of forecasting earnings is preferred for two reasons. First, the analysts' forecast errors are based on publicised forecasts made before earnings were announced. This in preference to time-series models either found ex post to provide the best fit as in Das *et al* (1998), or that were simply randomly chosen. Second, the evidence suggests that analysts provide more accurate forecasts than time-series models. If the market is aware of this analyst superiority then analysts' forecasts are more likely to reflect consensus

Table 6.1: **Explanatory Variables** – The variables are as presented in Table 3.2. In addition mean values and standard deviations are presented for all variables.

Explanatory Variable	Code	Mean Value	Std Dev
Earnings Yield	EY	0.0821	0.057
Dividend Yield	DY	0.0320	0.029
Market Capitalization	LNMC	8.086	1.310
Historical Earnings Growth	HGRO	0.130	0.125
Volatility of Past Earnings Growth	HVOL	0.277	0.291
Inflation	INF	0.077	0.021
Interest Rate	PRIME	0.1864	0.0214
Beta	BETA	0.842	0.364
Lagged Absolute Percentage Error	ABSPER_L	0.081	0.110
Lagged Signed Percentage Error	PERA_L	-0.030	0.133

market opinion of expected earnings than any individual time-series forecasting model.

Historical Earnings Volatility (HVOL)

HVOL was proposed by Albrecht *et al* (1977) and Foster (1986). If there is a link between the volatility of earnings changes and the difficulty of forecasting earnings, that relationship should be captured by this variable. This is a weaker variable than the lagged error in that it is an indirect measure of the unpredictability of earnings, and relies on a relationship between the volatility of earnings changes and the unpredictability of earnings. It is expected that this variable will be positively related to the absolute size of the error. It is not expected to be related to the forecast bias.

Table 6.2: **Balanced Panel Sample List** – The sample is made up of firms for which there is a forecast error in each of the years 1994-1999 as well as a full set of explanatory variables. Firms earnings less than 20 cents are excluded as are two firms with HVOL values more than four standard deviations from the mean.

Firm Name	TDC	Firm Name	TDC
AECI Ltd.	AFE	LTA Ltd.	LTA
African Oxygen Ltd	AFX	Nampak Ltd.	NPK
Allied Technologies Ltd	ALT	Palabora Mining Company Ltd.	PAM
Anglo American Plc.	AGL	Pepkor Ltd.	PEP
Anglo American Platinum Corporation Ltd.	AMS	Power Technologies Ltd	POW
Allied Electronics Corporation Ltd.	ATN	Pretoria Portland Cement Company Ltd.	PPC
De Beers	DBR	Richemont	RCH
Delta Electrical Industries Ltd.	DEL	Reunert Ltd.	RLO
Dorbyl Ltd.	DLV	Relyant	RLY
Edgars Consolidated Stores Ltd	ECO	Tiger Brands Ltd.	TBS
Ellerine Holdings Ltd.	ELH	Tongaat Hulett Group Ltd.	TNT
Fintech Ltd.	FIN	Trencor Ltd.	TRE
Foschini Ltd.	FOS	Unitrans Ltd	UTR
Highveld Steel and Vanadium Corporation Ltd.	HVL	Rembrandt Group Ltd	RMT
Impala Platinum Holdings Ltd.	IMP	Sappi Ltd.	SAP
JD Group Ltd.	JDG	Sun International (SA) Ltd.	SIS
Kersaf Investments Lt.	KER	Sasol	SOL

6.3.3 The balanced panel dataset

A balanced panel requires a full set of explanatory variables and dependent variables for all firms in the sample for all of the periods of the sample. The sample starts with the full 665 sample of forecast errors. Individual firm years were then eliminated if any explanatory variables were not available. In total 278 firm years were eliminated as a result of the non availability of explanatory variables. A further 29 firm years were eliminated where firms had earnings less than 20 cents. This was done to remove the skewing of results by small errors in cents per share resulting in large percentage errors. This left 357 firm years with a full set of explanatory variables.

The final choice in creating the balanced panel was in selecting sample length that would give the greatest number of firm years in the panel. A six year, 34-firm panel was chosen ahead of a seven year, 17-firm panel. A shorter period would not have added materially to the number of firms in the panel.

The list of explanatory variables, with brief summary statistics is presented in Table 6.1, the full panel list is presented in Table 6.2.

6.4 Methodology

A correlation matrix for all explanatory variables is presented in Table 6.3. As expected DY and EY are highly positively correlated.

The above studies, with the exception of Parkesh *et al* (1995), are based on pooled ordinary least squares regression, with adjustments to compensate for heteroskedasticity.

It is necessary to use a methodology that takes account of the fact that a panel of n firms for T years does not necessarily have the same information content as nT observations in one year. If the error term is independently and identically distributed in the time series and the cross section then it is appropriate to use

Table 6.3: **Explanatory Variable Correlation Matrix**

	EY	LNMC	DY	HGRO	HVOL	ABSPER_L	PERA_L	INF	PRIME	BETA
EY	1									
LNMC	-0.191	1								
DY	0.777	-0.193	1							
HGRO	0.194	0.030	0.018	1						
HVOL	0.197	-0.218	0.003	0.145	1					
ABSPER_L	-0.085	-0.033	-0.030	-0.046	0.130	1				
PERA_L	0.100	0.047	0.086	0.033	0.016	-0.537	1			
INF	-0.157	-0.032	-0.110	-0.156	0.012	-0.075	0.070	1		
PRIME	0.446	0.031	0.289	0.227	-0.016	-0.019	-0.014	-0.365	1	
BETA	-0.010	0.198	-0.072	-0.210	0.106	-0.054	0.023	-0.010	0.088	1

ordinary least squares regression. If this requirement is not met, two standard methods exist for analysing panel data, fixed effects and random effects regression.

If the intercept term is uncorrelated with the set of explanatory variables it is appropriate to use random effects regression. If the intercept term is correlated with the explanatory variables it is appropriate to use fixed effects regression.

Regression equations for each of the three methodologies are specified below.

6.4.1 Ordinary least squares regression

The simplest methodology for testing a problem of this type is to pool the data across years and perform an ordinary least squares regression. This was the basis of the studies by Kross *et al* (1990), Brown *et al* (1987) and Das *et al* (1998). All make some modifications to the traditional Ordinary Least Squares (OLS) regression model, for example to control for heteroskedasticity, but OLS remains the fundamental basis of their testing.

The OLS model takes the form:

$$y_{it} = \alpha + \sum_{k=1}^{K} \boldsymbol{\beta}_k \mathbf{X}_{kit} + u_{it}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T. \tag{6.3}$$

For the purposes of this study:

$$ABSPER_{it} = \alpha + \sum_{k=1}^{K} \boldsymbol{\beta}_k \mathbf{X}_{kit} + u_{it}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T, \tag{6.4}$$

$$PERA_{it} = \alpha + \sum_{k=1}^{K} \boldsymbol{\beta}_k \mathbf{X}_{kit} + u_{it}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T, \tag{6.5}$$

where y_{it} is the forecast error for firm i and period t, α and β_k are 1×1 and $1 \times k$ vectors of constants respectively; X_{kit} is a $k \times 1$ vector of exogenous variables and the error term, u_{it} is independently and identically distributed over i and t, with zero mean and variance σ_u^2.

An unmodified OLS model is the first model tested. The data for each of the six years are pooled and regressions are performed for the full sample as specified in (6.4) and (6.5).

6.4.2 Random effects panel regression

Two assumptions are required for the random effects model, the first is that individual effects α_i are random drawings from a common population. Second, it is required that the intercept terms are uncorrelated with the regressors. Given that there are clear and observable differences between firms it is unreasonable to assume that all firms are random drawings from the same population. Further, there is no reason to believe that the intercept terms are correlated with the explanatory variables.

The random effects model takes the form:

$$y_{it} = \sum_{k=1}^{K} \beta_k \mathbf{X}_{kit} + \epsilon_{it}, \quad \epsilon_{it} = \alpha_i + \eta_{it}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T, \qquad (6.6)$$

where α_i is uncorrelated with $\mathbf{X}_{kit}$.

For the purposes of this study:

$$ABSPER_{it} = \sum_{k=1}^{K} \beta_k \mathbf{X}_{kit} + \epsilon_{it}, \quad \epsilon_{it} = \alpha_i + \eta_{it}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T, \qquad (6.7)$$

$$PERA_{it} = \sum_{k=1}^{K} \beta_k \mathbf{X}_{kit} + \epsilon_{it}, \quad \epsilon_{it} = \alpha_i + \eta_{it}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T. \qquad (6.8)$$

The data for each of the six years are pooled and regressions are performed for the full sample as specified in (6.7) and (6.8).

6.4.3 Fixed effects panel regression

The first benefit of the fixed effects model is its theoretical ability to estimate coefficients efficiently where there are omitted variables. For a discussion of this the reader is directed to Johnston and DiNardo (1997:395–397) or Hsiao (2003:30). This has two implications. The first is efficient coefficient estimates for included variables where there are omitted variables. The second is that it is not possible to capture the effect of time invariant factors. In the context of this study it is not possible to capture the effect that the industry to which a firm belongs has on the error because the industry classification is time invariant.

The second benefit is that the OLS assumption of a common intercept for all firms for all periods is relaxed. The most common form of fixed effects model is the one estimated here. The use of a fixed effect panel model with cross sectional dummies estimates an intercept term for each firm, and the slope coefficients after the effect of the different intercepts has been removed for. Given the nature of the data this model is particularly appealing. Differently risky firms may then have different unique errors, based on the riskiness of the underlying business, and still respond to the treatment variables in a similar fashion to other firms.

The cost of estimating unique intercepts for each firm is the loss of degrees of freedom.

The model then takes the following general form:

$$y_{it} = \alpha_i + \sum_{k=1}^{K} \beta_k X_{kit} + u_{it}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T. \tag{6.9}$$

For the purposes of this study:

$$ABSPER_{it} = \alpha_i + \sum_{k=1}^{K} \beta_k X_{kit} + u_{it}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T, \tag{6.10}$$

$$PERA_{it} = \alpha_i + \sum_{k=1}^{K} \beta_k X_{kit} + u_{it}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T. \tag{6.11}$$

All variables are as defined previously. The significant change is that α is now estimated for each individual firm. This model differs from the random effects model in that α_i is assumed to be correlated with $\mathbf{X}_{kit}$.

The data for each of the six years are pooled and regressions are performed for the full sample as specified in (6.10) and (6.11).

6.5 Results

Given that this study is concerned with firms with unique characteristics that effect both the error term and are related to the explanatory variables, the emphasis of the study is on the fixed effects model. Further, if the random effects model is valid, 'the fixed effects estimator will still produce consistent estimates of identifiable parameters' Johnston and DiNardo (1997:403). The results for the ordinary least squares regression and the random effects regression are discussed briefly, with the regression output presented in Appendix B. The results for the fixed effects regression output for $PERA$ and $ABSPER$ are presented in Tables 6.4 and 6.5 respectively.

6.5.1 Signed percentage error

The pooled OLS model provides a poor fit. None of the variables are significant, and the adjusted R^2 is less than 1%. Similarly, the random effects model provides a poor fit.

The fixed effects model provides a better fit. Only the lagged error term is significant. The negative slope coefficient suggests that the direction of analyst bias changes, supporting the hypothesis that analysts learn from their past mistakes. As expected, given the mean bias for the full sample, the intercepts are uniformly less than zero.

Overall the results are consistent with the expectation that the fixed effects model is appropriate to the data in question. If individual firms are treated as uniquely risky, and thus allowed a unique intercept term, a statistically significant relationship is found between the analysts' forecast bias and the forecast bias in the previous period. The implication of this is that an optimistic (pessimistic) forecast is likely to be followed by less optimistic (pessimistic) forecast or a pessimistic (optimistic) forecast.

6.5.2 Absolute percentage error

For the pooled OLS the lagged error term is the only significant variable. The R^2 is very low at 2.6%. Again, the random effects model provides a poor fit.

The fixed effects model provides a significantly better fit; the lagged error term and firm size are related to the absolute size of the error. The direction of the estimated slope coefficient for the size factor is as expected - larger (smaller) firms are associated with smaller (larger) absolute percentage errors. The lagged error term is particularly interesting given the nature of the data. The sign of the slope coefficient suggests a negative relationship between the past error and the current error. This suggests negative serial correlation in the forecasts. An error larger than the mean in the previous period is expected to be followed by a smaller error in the current period. An error smaller than the mean in the previous period is expected to be followed by a larger error in the current period.

Overall the fixed effects model provided the best fit. The results are consistent with the prior literature on asset pricing - that larger firms are associated with smaller forecast errors. The greater certainty attached to the forecast earnings of large companies suggests a lower expected return. The surprising result, and important one for the question of rational expectations is that the forecast error in any period is negatively related to the error in the previous period.

Table 6.4: **Balanced Panel Fixed Effects Results for the Signed Percentage Error** – The regression $PERA_{it} = \alpha_i + \sum_{k=1}^{K} \beta_k \mathbf{X}_{kit} + u_{it}$ is estimated for the 34 firm balanced panel for the period 1994–1999.

Variable	Slope Coefficient	p Value
EY	0.0375	0.9436
LNMC	0.0553	0.2182
DY	-0.1648	0.8836
HGRO	-0.0546	0.6803
HVOL	-0.0098	0.9053
PERA_L	-0.2592	**0.0156**
INF	-0.5693	0.3844
PRIME	-0.1337	0.8737
BETA	-0.0892	0.2158

Firm	Intercept	Firm	Intercept	Firm	Intercept
AFE	-0.340279	FOS	-0.464198	RCH	-0.367292
AFX	-0.421085	HVL	-0.231178	RLO	-0.257914
AGL	-0.3936	IMP	-0.2131	RLY	-0.1586
ALT	-0.2805	JDG	-0.3358	RMT	-0.3821
AMS	-0.4778	KER	-0.3047	SAP	-0.2533
ATN	-0.2949	LTA	-0.1608	SIS	-0.3532
DBR	-0.4723	NPK	-0.3122	SOL	-0.5136
DEL	-0.1968	PAM	-0.1587	TBS	-0.3509
DLV	-0.4949	PEP	-0.3403	TNT	-0.3413
ECO	-0.3596	POW	-0.2815	TRE	-0.7120
ELH	-0.2072	PPC	-0.2984	UTR	-0.5035
FIN	-0.196				

R-squared	0.2812	Mean dependent var	-0.0369
Adjusted R-squared	0.0937	S.D. dependent var	0.1837
S.E. of regression	0.1749	Sum squared resid	4.9234
F-statistic	7.8748	Durbin-Watson stat	1.8740
Prob(F-statistic)	0.0000		

Table 6.5: **Balanced Panel Fixed Effects Results for the Absolute Percentage Error** – The regression $ABSPER_{it} = \alpha_i + \sum_{k=1}^{K} \beta_k \mathbf{X}_{kit} + u_{it}$ is estimated for the 34 firm balanced panel for the period 1994–1999.

	Variable		Slope Coefficient		p Value
	EY		0.1129		0.7867
	LNMC		-0.0771		**0.0291**
	DY		-0.5200		0.5543
	HGRO		-0.0005		0.9961
	HVOL		-0.0262		0.6865
	ABSPER_L		-0.2982		**0.0091**
	INF		0.2346		0.6482
	PRIME		0.8342		0.2061
	BETA		-0.0131		0.8155
Firm	**Intercept**	**Firm**	**Intercept**	**Firm**	**Intercept**
AFE	0.5979	FOS	0.6207	RCH	0.7453
AFX	0.6110	HVL	0.5669	RLO	0.5090
AGL	0.7567	IMP	0.6581	RLY	0.3274
ALT	0.4016	JDG	0.5571	RMT	0.6449
AMS	0.8850	KER	0.5073	SAP	0.6850
ATN	0.5063	LTA	0.3480	SIS	0.5766
DBR	0.8021	NPK	0.5704	SOL	0.6644
DEL	0.4111	PAM	0.5870	TBS	0.5917
DLV	0.8982	PEP	0.6676	TNT	0.5685
ECO	0.5851	POW	0.5152	TRE	0.8877
ELH	0.4913	PPC	0.5367	UTR	0.7032
FIN	0.3789				
R-squared		0.4105		Mean dependent var	0.0991
Adjusted R-squared		0.2567		S.D. dependent var	0.1589
S.E. of regression		0.1370		Sum squared resid	3.0197
F-statistic		14.0137		Durbin-Watson stat	1.9005
Prob(F-statistic)		0.0000			

Out of sample testing

The PERA regression results are used to try and improve the analysts' forecasts. The methodology and results are contained in Appendix C. While potentially unreliable (data selection problems are discussed in Appendix C), the notable finding is that the forecasts cannot be improved without increasing the standard deviation of the forecast errors.

6.6 Conclusions

Signed percentage errors are inversely related to errors made in the previous period. This may be explained by analysts learning from past mistakes (biases) and correcting, sometimes overcorrecting, for them in the following period.

Absolute percentage errors are inversely related to past errors and the firm size. A cost to information argument could be made for the firm size finding - larger firms are more important to money managers as they form a larger part of the benchmark against which their performance is measured. They then have good reason to expend greater resources on forecasting earnings for these large firms than for small firms. Given this additional effort it is then not surprising that large firms are associated with smaller forecast errors. The finding on the negative serial correlation in the errors is more surprising.

Chapter 7

Forecast Predictability – An Unbalanced Panel Approach

7.1 Introduction

The panel analysis of Chapter 6 is expanded to include firms with fewer than six years of observations, or observations that do not fall into the 1994–1999 period. The cost of increasing the sample is that for the fixed and random effects models the intercept estimates are less reliable, as they are estimated with fewer observations per firm. The first advantage of expanding the sample is that more observations are available for estimating the slope coefficients. The second advantage is that it makes it possible to estimate the models using only the Industrial group.

7.2 Data

The explanatory variables are the same as those used for the balanced panel. The only change to the selection criteria for inclusion is that firms must have at least four years of explanatory variables in order to be included. The four years do not need to be consecutive. This led to two changes to the panel. First, additional observations were added to firms from the balanced panel in the 1991–1993 period. Second, eight firms were added to the sample. All eight firms added fall into the broad industry classification as Industrials. The result is a sample with a total of 277 firm years.

Bias may arise if the missing observations are other than randomly missing. It does not appear as though any of the missing data are missing for any systematic reasons. It is therefore expected that the unbalanced panel results will not be biased by the missing observations.

Table 7.1: **Unbalanced Panel Sample List** – For the unbalanced panel firms with at least four years of explanatory variables are included. The observations do not need to be in consecutive years. The remaining selection criteria are the same as for the balanced panel. Firms not included in the balanced panel are marked with a (*) after the name.

Firm Name	**Years in Sample**	**TDC**
Amalgamated Beverage Industries Ltd *	1995-1999	ABI
AECI Ltd	1991-1999	AFE
African Oxygen Ltd	1991-1999	AFX
Allied Technologies Ltd	1994-1999	ALT
Anglo American Plc	1993-1999	AGL
Anglo American Platinum Corp Ltd.	1993-1999	AMS
Allied Electronics Corporation Ltd	1994-1999	ATN
Cadbury Schweppes (SA) Ltd *	1996-1999	CAS
De Beers	1993-1999	DBR
Delta Electrical Industries Ltd	1994-1999	DEL
Dorbyl Ltd.	1994-1999	DLV
Didata *	1996-1999	DDT

Table 7.1: **Unbalanced Panel Sample List** – For the unbalanced panel firms with at least four years of explanatory variables are included. The observations do not need to be in consecutive years. The remaining selection criteria are the same as for the balanced panel. Firms not included in the balanced panel are marked with a (*) after the name.

Edgars Consolidated Stores Ltd	1991-1999	ECO
Ellerine Holdings Ltd	1994-1999	ELH
Fintech Ltd	1994-1999	FIN
Foschini Ltd	1994-1999	FOS
Highveld Steel and Vanadium Corp Ltd	1994-1999	HVL
Hudaco Industries Ltd *	1994-1998	HDC
Impala Platinum Holdings Ltd.	1993-1999	IMP
JD Group Ltd	1994-1999	JDG
Kersaf Investments Ltd	1993-1999	KER
LTA Ltd	1994-1999	LTA
Medi-Clinic Corporation Ltd *	1996-1999	MDC
Murray and Roberts Holdings Ltd *	1991-1996	MUR
Nampak Ltd	1993-1999	NPK
Palabora Mining Company Ltd	1994-1999	PAM
Pepkor Ltd	1991-1999	PEP
Pick n Pay Stores Ltd *	1994,1996-1999	PIK
Power Technologies Ltd	1994-1999	POW
Pretoria Portland Cement Company Ltd	1993-1999	PPC
Richemont	1994-1999	RCH
Rembrandt Group Ltd	1994-1999	RMT
Reunert Ltd	1991-1999	RLO
Relyant Retail Ltd	1995-1999	RLY
Sappi Ltd.	1991-1999	SAP

Table 7.1: **Unbalanced Panel Sample List** – For the unbalanced panel firms with at least four years of explanatory variables are included. The observations do not need to be in consecutive years. The remaining selection criteria are the same as for the balanced panel. Firms not included in the balanced panel are marked with a (*) after the name.

Sasol	1992-1999	SOL
Shoprite Holdings Ltd *	1995-1999	SHP
Sun International (SA) Ltd	1993-1999	SIS
Tiger Brands Ltd	1991-1999	TBS
Tongaat Hulett Group Ltd	1993-1999	TNT
Trencor Ltd.	1994-1999	TRE
Unitrans Ltd	1994-1999	UTR
Wooltru Ltd *	1994-1997	WLO

Table 7.1 contains a full list of all firms in the unbalanced panel, as well as the years for which observations were available for the firm. Firms that were not included in the balanced panel are marked with a (*) after the firm name.

7.3 Methodology and results

A correlation matrix for the explanatory variables is presented in Table 7.2.

The methodology for the Ordinary Least Squares, Fixed Effects and Random Effects regressions are the same as in the previous chapter. The results for the OLS and Random Effects regressions are presented in Appendix D.

The results for PERA, presented in Table 7.3, are similar to those for the balanced panel. The only significant variable is the lagged error term. The slope coefficient is again negative, indicating that analysts learn from previous errors and reduce the bias in the direction of the previous bias. The intercept terms are

	EY	LNMC	DY	HGRO	HVOL	ABSPER_L	PERA_L	INF	PRIME	BETA
EY	1									
LNMC	-0.259	1								
DY	0.797	-0.235	1							
HGRO	0.022	0.043	-0.046	1						
HVOL	0.133	-0.173	-0.011	0.167	1					
ABSPER_L	-0.100	-0.033	-0.054	-0.091	0.129	1				
PERA_L	0.055	0.059	0.064	0.062	-0.028	-0.464	1			
INF	-0.064	-0.075	-0.013	0.057	0.216	-0.081	0.091	1		
PRIME	0.385	-0.010	0.246	0.218	0.026	-0.050	-0.015	-0.160	1	
BETA	0.035	0.211	-0.016	-0.181	-0.004	-0.051	0.014	-0.086	0.054	1

Table 7.2: **Explanatory Variable Correlation Matrix**

Table 7.3: **Unbalanced Panel Fixed Effects Results for the Signed Percentage Error** – The regression $PERA_{it} = \alpha_i + \sum_{k=1}^{K} \beta_k \mathbf{X}_{kit} + u_{it}$ is estimated for the 43 firm unbalanced panel. Firms with at least four years of observations are included, the four years do not need to be consecutive.

Variable	Slope Coefficient	p Value
EY	0.0453	0.9298
LNMC	0.0473	0.2616
DY	-0.0486	0.9626
HGRO	-0.0327	0.7172
HVOL	0.0920	0.1070
PERA_L	-0.2041	**0.0401**
INF	0.1920	0.7221
PRIME	-0.2859	0.7037
BETA	-0.0958	0.1241

Firm	Intercept	Firm	Intercept	Firm	Intercept
ABI	-0.2525	FOS	-0.4489	PPC	-0.2740
AFE	-0.3008	HDC	-0.2252	RCH	-0.3368
AFX	-0.3940	HVL	-0.2810	RLO	-0.2402
AGL	-0.3248	IMP	-0.1937	RLY	-0.2512
ALT	-0.2674	JDG	-0.3210	RMT	-0.3422
AMS	-0.4142	KER	-0.2867	SAP	-0.3446
ATN	-0.2853	LTA	-0.1662	SHP	-0.7555
CAS	-0.2803	MDC	-0.3920	SIS	-0.3313
DBR	-0.4443	MUR	-0.3955	SOL	-0.4645
DDT	-0.2959	NPK	-0.2815	TBS	-0.3132
DEL	-0.1976	PAM	-0.1971	TNT	-0.2853
DLV	-0.5300	PEP	-0.3689	TRE	-0.6999
ECO	-0.3212	PIK	-0.2271	UTR	-0.4921
ELH	-0.1921	POW	-0.2693	WLO	-0.2769
FIN	-0.2397				

R-squared	0.2553	Mean dependent var	-0.0352
Adjusted R-squared	0.0865	S.D. dependent var	0.1962
S.E. of regression	0.1875	Sum squared resid	7.9082
F-statistic	9.6403	Durbin-Watson stat	1.8734
Prob(F-statistic)	0.0000	Total Observations	277

again uniformly negative. The adjusted R-squared is again less than 10%.

The results for ABSPER, presented in Table 7.4 are more interesting. The firm size remains significant, the coefficient remains negative, and the absolute value of the coefficient is larger.

The lagged error term coefficient remains negative, with an absolute value slightly smaller than for the balance panel. It is however no longer significant at the 5% level (p = 0.055). HVOL is now significant (p = 0.001), the slope coefficient is negative, with a larger absolute value than for the balanced panel. The adjusted R^2 is lower than that of the balanced panel but remains in excess of 20%.

The ABSPER regression is then re-estimated omitting HVOL and ABSPER_L in turn, the results of these regressions are presented in Table 7.5. The results suggest that despite their low correlation the two variables capture the same underlying process, and that HVOL captures the process more completely than ABSPER_L. The addition of ABSPER_L adds explanatory power to the model after the inclusion, however the incremental value is small (the adjusted R^2 increases by 0.009).

7.4 Industry differences

A shortcoming of the fixed effects regression is that it is not possible to capture the effect of time invariant variables. In this analysis the most significant omission is that the industry to which a firm belongs has not been considered as a candidate variable to explain the differences in forecasting errors between firms. Given the enlarged sample available using an unbalanced panel there are sufficient observations, both individual firms and firms years, for Industrial firms to estimate the regressions using only firms from this sector. The results of the Fixed Effects regression are presented in Tables 7.6 and 7.7 respectively.

The results for ABSPER are similar in finding a negative relationship be-

Table 7.4: **Unbalanced Panel Fixed Effects Output for the Absolute Percentage Error** – The regression $ABSPER_{it} = \alpha_i + \sum_{k=1}^{K} \beta_k \mathbf{X}_{kit} + u_{it}$ is estimated for the 43 firm unbalanced panel. Firms with at least four years of observations are included, the four years do not need to be consecutive.

Variable	Slope Coefficient	p Value
EY	-0.1222	0.7723
LNMC	-0.1108	**0.0014**
DY	-0.8736	0.3011
HGRO	0.0188	0.7997
HVOL	-0.1596	**0.0007**
ABSPER_L	-0.2188	0.055
INF	-0.6360	0.1514
PRIME	1.0171	0.0974
BETA	0.0300	0.5541

Firm	**Intercept**	**Firm**	**Intercept**	**Firm**	**Intercept**
ABI	0.8364	FOS	0.9359	PPC	0.8302
AFE	0.9314	HDC	0.6464	RCH	1.1298
AFX	0.9185	HVL	0.9330	RLO	0.7784
AGL	1.1124	IMP	0.9789	RLY	0.7185
ALT	0.6949	JDG	0.8476	RMT	1.0050
AMS	1.1951	KER	0.8311	SAP	1.094
ATN	0.7776	LTA	0.6066	SHP	1.3301
CAS	0.7354	MDC	0.8090	SIS	0.9057
DBR	1.2030	MUR	1.1494	SOL	1.0595
DDT	0.9461	NPK	0.8894	TBS	0.9078
DEL	0.6751	PAM	0.9966	TNT	0.9130
DLV	1.2231	PEP	0.9973	TRE	1.2040
ECO	0.8945	PIK	0.6880	UTR	0.9771
ELH	0.7611	POW	0.7900	WLO	0.9351
FIN	0.7036				

R-squared	0.3696	Mean dependent var	0.0977
Adjusted R-squared	0.2267	S.D. dependent var	0.1736
S.E. of regression	0.1527	Sum squared resid	5.2439
F-statistic	16.4866	Durbin-Watson stat	1.8117
Prob(F-statistic)	0.0000	Total Observations	277

Table 7.5: **Unbalanced Panel Regression Omitting HVOL and ABSPER_L** – The regression $ABSPER_{it} = \alpha_i + \sum_{k=1}^{K} \beta_k \mathbf{X}_{kit} + u_{it}$ is re-estimated, omitting HVOL and ABSPER_L in turn.

Omitting HVOL		**Omitting ABSPER_L**	
ABSPER_L	-0.2692	**HVOL**	-0.1711
	0.0202		**0.0003**
EY	-0.1990	**EY**	-0.0020
	0.6451		0.9961
LNMC	-0.0781	**LNMC**	-0.1069
	0.0214		**0.0021**
DY	-0.2747	**DY**	-0.9549
	0.7453		0.2607
HGRO	-0.0150	**HGRO**	0.0377
	0.8421		0.6092
INF	-0.7882	**INF**	-0.5051
	0.0813		0.2513
PRIME	0.7577	**PRIME**	0.9766
	0.2236		0.1133
BETA	0.0365	**BETA**	0.0355
	0.4821		0.4863
R-squared	0.3362	R-squared	0.3591
Adjusted R-squared	0.1893	Adjusted R-squared	0.2174
F-statistic	16.3484	F-statistic	18.0923
Durbin-Watson stat	1.7613	Durbin-Watson stat	1.9368

Table 7.6: Unbalanced Panel Fixed Effects Regression Output for the Absolute Percentage Error for Industrial Firms Only– The regression $PERA_{it} = \alpha_i + \sum_{k=1}^{K} \beta_k \mathbf{X}_{kit} + u_{it}$ is estimated for the 33 Industrial firms in the unbalanced panel. Firms with at least four years of observations are included, the four years do not need to be consecutive.

	Variable		**Slope Coefficient**		**p Value**
	EY		-0.9927		0.1807
	LNMC		-0.1999		**0.0001**
	DY		-0.2610		0.8749
	HGRO		0.0340		0.7181
	HVOL		-0.2340		**0.0001**
	ABSPER_L		-0.2458		0.0877
	INF		-0.8552		0.1491
	PRIME		1.4961		0.0626
	BETA		0.0822		0.2190
Firm	**Intercept**	**Firm**	**Intercept**	**Firm**	**Intercept**
ABI	1.4898	HDC	1.1899	RCH	2.0005
ALT	1.2754	JDG	1.4689	RLO	1.3907
ATN	1.3429	KER	1.4899	RLY	1.2875
CAS	1.3467	LTA	1.1155	RMT	1.8287
DDT	1.6811	MDC	1.4060	SHP	2.0473
DEL	1.1937	MUR	1.9045	SIS	1.5835
DLV	1.8458	NPK	1.6100	TBS	1.6410
ECO	1.5759	PEP	1.6829	TNT	1.6107
ELH	1.3491	PIK	1.2253	TRE	1.8750
FIN	1.2671	POW	1.3751	UTR	1.5550
FOS	1.6242	PPC	1.4616	WLO	1.6032
R-squared		0.4184	Mean dependent var		0.0967
Adjusted R-squared		0.2693	S.D. dependent var		0.1952
S.E. of regression		0.1668	Sum squared resid		4.4524
F-statistic		14.3849	Durbin-Watson stat		1.8037
Prob(F-statistic)		0.0000	Total Observations		202

Table 7.7: **Unbalanced Panel Fixed Effects Regression Output for the Signed Percentage Error for Industrial Firms Only**– The regression $ABSPER_{it} = \alpha_i + \sum_{k=1}^{K} \beta_k \mathbf{X}_{kit} + u_{it}$ is estimated for the 33 firms in the unbalanced panel. Firms with at least four years of observations are included, the four years do not need to be consecutive.

Variable	Slope Coefficient	p Value
EY	0.1600	0.8559
LNMC	0.1139	0.0548
DY	1.1388	0.5703
HGRO	-0.0381	0.7344
HVOL	0.1378	0.0575
PERA_L	-0.2726	**0.0320**
INF	0.4002	0.5761
PRIME	-1.5044	0.1212
BETA	-0.0833	0.3034

Firm	Intercept	Firm	Intercept	Firm	Intercept
ABI	-0.6136	HDC	-0.5175	RCH	-0.8468
ALT	-0.5836	JDG	-0.6800	RLO	-0.6039
ATN	-0.5861	KER	-0.69362	RLY	-0.5188
CAS	-0.5966	LTA	-0.4442	RMT	-0.8356
DDT	-0.6832	MDC	-0.6842	SHP	-1.1393
DEL	-0.4738	MUR	-0.8308	SIS	-0.7787
DLV	-0.8827	NPK	-0.7151	TBS	-0.7448
ECO	-0.7188	PEP	-0.7534	TNT	-0.6884
ELH	-0.5311	PIK	-0.5172	TRE	-1.0732
FIN	-0.5426	POW	-0.5900	UTR	-0.8071
FOS	-0.8180	PPC	-0.6730	WLO	-0.6624

R-squared	0.2866	Mean dependent var	-0.0436
Adjusted R-squared	0.1037	S.D. dependent var	0.2135
S.E. of regression	0.2021	Sum squared resid	6.5350
F-statistic	8.0329	Durbin-Watson stat	1.8255
Prob(F-statistic)	0.0000	Total Observations	202

tween firm size and the size of the forecast error. The results for HVOL and ABSPER_L mimic those of the larger unbalanced panel. For PERA, only the lagged error term is significant at the 5% level, and in the same direction as previously. The surprise is that the firm size and historical volatility of earnings changes are significant (p=0.0548 and p=0.0575 respectively), both with positive slope coefficients. This suggests that larger (smaller) firms and firms with more (less) volatile earnings changes are associated with relatively optimistically pessimistically) biased forecasts.

7.5 Conclusions

Expanding the sample size produced similar results to those found with the balanced panel for the signed percentage error. The lagged error term (PERA_L) remains the only significant variable. The results for the absolute percentage error are different in that the volatility of past earnings is significant. Omitting the volatility of past earnings changes (HVOL) and the lagged error term (ABSPER_L) in turn suggested that the two variables capture the same underlying process, and that for the sample the volatility of past earnings changes largely subsumed the lagged error term. Given the increased sample size it was also possible to estimate the model for Industrial firms only. The results were largely the same as for the full unbalanced panel. This is not surprising for two reasons. First, the Industrial grouping is the largest part of the full unbalanced panel to which it is being compared. Second, it confirms the finding that the largest part of the forecast error is made at the firm rather than the industry or full sample level. The historical volatility of earnings changes and firm size were significant for the absolute percentage error at the 5% significance level, while the lagged error term is significant at the 10% significance level. The results for the signed percentage error were similar in that the lagged error term is significant at the 5% significance level, what is new, and unexpected, is that the firm size and past earnings change volatility are significant at the 10% level.

Chapter 8

Conclusions

This thesis considers the question of whether analysts' forecasts represent rational expectations for company earnings. The question is divided into three parts, with each part considered separately. The first part deals with the bias in the analysts' forecasts, the second with whether the analysts' forecasts are the most accurate available, and the third with the question of patterns in the forecast errors.

The study is based on a sample of 665 earnings forecasts taken from the I/B/E/S database. The study includes all JSE Securities Exchange listed firms for which forecasts were available during the period January 1990 to December 1999.

In response to the question of bias in the analysts' forecasts, weak evidence of an optimistic bias is found. While the intercepts of the regression of the actual change on the forecast change are consistently negative (indicating an optimistic bias), the intercepts are not significantly different from zero for the full sample or any of the individual years of the study. The median percentage error is, however, zero. Decomposing the mean squared forecast error suggests that the largest part of the error is made at the individual firm level rather than the industry or full sample level.

On the question of analysts' forecasts as the source of the most accurate forecasts available, the analysts' forecasts are compared to forecasts produced by nine econometric models based on extrapolations of past earnings. The analysts produced forecasts that are more accurate than those produced by the time-series extrapolations of past earnings. This finding is confirmed by the Thiel's U statistics as well as the results of the Wilcoxon Signed Ranks test. Two of the time-series models produced less biased forecasts, however the mean absolute errors are smaller for analysts than any of the models tested. The analysts' advantage over the time-series forecasts increases as the earnings announcement date approaches.

On the final question of patterns in the errors, the bias is found to be inversely related to the forecast bias in the previous period. The magnitude of the error is found to be inversely related to the firm size and the absolute percentage error from the previous period.

Overall the forecasts do not meet the strict requirements for rational expectations. The findings suggest that analysts did not make full use of all publicly available information in producing the forecasts.

Suggestions for future research are (i) to test the extent to which the forecast error explains abnormal returns after earnings are announced and (ii) to apply the methodology of Chapter Six of this study using additional information about the forecasts, specifically the number of analysts and the dispersion of the analysts' forecasts.

Appendix A

Appendix to Chapter 4 – Forecast errors as the earnings announcement date approaches

Table A.1: **The Absolute Percentage Error the as Earnings Announcement Date Approaches** – The forecast in each month is compared to the realised earnings, and the absolute percentage error is calculated as $|(A_t - F_{t-n})/A_t|$ for $n = 1$ to 12 for each monthly observation.

	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1
1990			**157%**	**119%**	**89%**	**72%**	**36%**	**28%**	**50%**	**47%**	**4%**	**17%**
n		1	6	8	11	15	17	22	32	35	42	43
1991	**17%**	**14%**	**14%**	**13%**	**10%**	**12%**	**13%**	**10%**	**10%**	**11%**	**10%**	**8%**
n	30	47	49	49	49	49	49	49	49	49	49	50
1992	**16%**	**17%**	**17%**	**12%**	**14%**	**13%**	**13%**	**15%**	**16%**	**16%**	**16%**	**15%**
n	42	48	50	50	50	50	50	51	51	51	51	51
1993	**35%**	**21%**	**18%**	**12%**	**12%**	**12%**	**10%**	**10%**	**11%**	**11%**	**11%**	**9%**
n	47	50	50	50	50	50	51	51	51	51	51	51
1994	**25%**	**19%**	**32%**	**28%**	**28%**	**26%**	**27%**	**15%**	**16%**	**15%**	**15%**	**13%**
n	49	52	52	52	53	53	53	54	54	54	54	54
1995	**25%**	**24%**	**17%**	**18%**	**17%**	**18%**	**19%**	**19%**	**20%**	**13%**	**14%**	**12%**
n	52	57	57	57	57	58	58	58	58	58	58	58
1996	**22%**	**20%**	**16%**	**14%**	**17%**	**13%**	**14%**	**14%**	**15%**	**15%**	**14%**	**16%**
n	62	68	68	68	68	68	68	68	68	68	68	69
1997	**41%**	**37%**	**40%**	**39%**	**25%**	**26%**	**26%**	**25%**	**25%**	**22%**	**20%**	**19%**
n	69	73	76	77	77	77	77	78	78	78	78	78
1998	**70%**	**38%**	**45%**	**31%**	**28%**	**24%**	**20%**	**19%**	**23%**	**16%**	**14%**	**14%**
n	89	94	94	95	96	97	97	98	98	98	98	98
1999	**107%**	**101%**	**96%**	**90%**	**70%**	**64%**	**54%**	**36%**	**37%**	**47%**	**46%**	**45%**
n	96	102	107	108	108	110	110	111	111	111	111	113

Table A.2: **The Signed Percentage Error as the Earnings Announcement Date Approaches** – The forecast in each month is compared to the realised earnings, and the absolute percentage error is calculated as $(A_t - F_{t-n})/A_t$ for $n = 1$ to 12 for each monthly observation.

	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1
1990			**-153%**	**-116%**	**-86%**	**-56%**	**-21%**	**-15%**	**19%**	**17%**	**14%**	**-11%**
n		1	6	8	11	15	17	22	32	35	42	43
1991	**7%**	**4%**	**3%**	**2%**	**3%**	**0%**	**1%**	**3%**	**1%**	**-2%**	**-1%**	**-3%**
n	30	47	49	49	49	49	49	49	49	49	49	50
1992	**-4%**	**-9%**	**-8%**	**-3%**	**-5%**	**-3%**	**-4%**	**-4%**	**-4%**	**-4%**	**-4%**	**-7%**
n	42	48	50	50	50	50	50	51	51	51	51	51
1993	**-26%**	**-14%**	**-10%**	**-4%**	**-4%**	**-3%**	**-1%**	**-1%**	**-1%**	**-1%**	**-1%**	**-3%**
n	47	50	50	50	50	50	51	51	51	51	51	51
1994	**-4%**	**-5%**	**-17%**	**-15%**	**-16%**	**-15%**	**-15%**	**-7%**	**-7%**	**-4%**	**-4%**	**-5%**
n	49	52	52	52	53	53	53	54	54	54	54	54
1995	**-3%**	**-12%**	**-6%**	**-5%**	**-5%**	**-6%**	**-9%**	**-12%**	**-12%**	**-5%**	**-6%**	**-8%**
n	52	57	57	57	57	58	58	58	58	58	58	58
1996	**-3%**	**-6%**	**-1%**	**1%**	**-4%**	**0%**	**-2%**	**-3%**	**-4%**	**-4%**	**-4%**	**-5%**
n	62	68	68	68	68	68	68	68	68	68	68	69
1997	**5%**	**-28%**	**-34%**	**-32%**	**-18%**	**-19%**	**-18%**	**-18%**	**-18%**	**-16%**	**-13%**	**-14%**
n	69	73	76	77	77	77	77	78	78	78	78	78
1998	**-47%**	**-21%**	**-28%**	**-14%**	**-11%**	**-7%**	**-6%**	**-5%**	**1%**	**-6%**	**-1%**	**0%**
n	89	94	94	95	96	97	97	98	98	98	98	98
1999	**-76%**	**-74%**	**-70%**	**-67%**	**-53%**	**-48%**	**-43%**	**-12%**	**-12%**	**-36%**	**-37%**	**-37%**
n	96	102	107	108	108	110	110	111	111	111	111	113

Appendix B

Appendix to Chapter 6a – Ordinary least squares and random effects regression output for the balanced panel

Table B.1: **Ordinary Least Squares Regression Output for the Signed Percentage Error** – A balanced panel is constructed with all firms with a full set of explanatory variables for the period 1994-1999. Firms reporting less than 20 cents in earnings are eliminated from the sample, as are two firms with HVOL values more than four standard deviations from the mean. The result is a 34 firm panel. The following model is then estimated: $PERA_{it} = \alpha + \sum_{k=1}^{K} \beta_k \mathbf{X}_{kit} + u_{it}$.

Variable		Coefficient	p Value
Intercept		-0.0500	0.7774
EY		-0.2098	0.6494
LNMC		0.0122	0.2598
DY		0.5023	0.5106
HGRO		-0.0557	0.6320
HVOL		0.0304	0.5435
PERA_L		0.0298	0.7646
INF		-0.5907	0.3825
PRIME		-0.2393	0.7503
BETA		0.0061	0.8764
R-squared	0.0155	Mean dependent var	-0.0369
Adjusted R-squared	-0.0302	S.D. dependent var	0.1837
S.E. of regression	0.1865	Sum squared resid	6.7441
F-statistic	0.3383	Durbin-Watson stat	1.7034
Prob(F-statistic)	0.9612		

Table B.2: **Ordinary Least Squares Regression Output for the Absolute Percentage Error**– A balanced panel is constructed with all firms with a full set of explanatory variables for the period 1994-1999. Firms reporting less than 20 cents in earnings are eliminated from the sample, as are two firms with HVOL values more than four standard deviations from the mean. The result is a 34 firm panel. The following model is then estimated: $ABSPER_{it} = \alpha + \sum_{k=1}^{K} \beta_k \mathbf{X}_{kit} + u_{it}$.

Variable		**Coefficient**	**p Value**
Intercept		-0.0528	0.725
EY		0.3100	0.432
LNMC		-0.0059	0.5237
DY		-0.2897	0.6553
HGRO		0.0077	0.9376
HVOL		0.0107	0.8038
ABSPER_L		0.2620	**0.0128**
INF		0.6968	0.2254
PRIME		0.6384	0.3166
BETA		-0.0178	0.5915
R-squared	0.0543	Mean dependent var	0.0991
Adjusted R-squared	0.0104	S.D. dependent var	0.1589
S.E. of regression	0.1580	Sum squared resid	4.8441
F-statistic	1.2381	Durbin-Watson stat	1.7422
Prob(F-statistic)	0.2738		

Table B.3: **Balanced Panel Random Effects Regression Output for the Signed Percentage Error** – A balanced panel is constructed with all firms with a full set of explanatory variables for the period 1994–1999. Firms reporting less than 20 cents in earnings are eliminated from the sample, as are two firms with HVOL values more than four standard deviations from the mean. The result is a 34 firm panel. The following model is then estimated: $PERA_{it} = \sum_{k=1}^{K} \beta_k \mathbf{X}_{kit} + \epsilon_{it}, \quad \epsilon_{it} = \alpha_i + \eta_{it}$

Variable	Coefficient	p Value
Intercept	-0.0408	0.8175
EY	-0.2250	0.6250
LNMC	0.0111	0.2697
DY	0.5697	0.4446
HGRO	-0.0595	0.6076
HVOL	0.0309	0.5158
PERA_L	0.0810	0.4168
INF	-0.5947	0.3894
PRIME	-0.2564	0.7371
BETA	0.0116	0.7530

Firm	Intercept	Firm	Intercept	Firm	Intercept
AFE	0.0015	FOS	0.0154	RCH	-0.0212
AFX	0.0083	HVL	-0.0105	RLO	-0.0059
AGL	0.0037	IMP	-0.0173	RLY	-0.0164
ALT	-0.0121	JDG	0.0037	RMT	-0.0088
AMS	0.0220	KER	-0.00722	SAP	-0.0127
ATN	-0.0035	LTA	-0.01878	SIS	0.0119
DBR	0.0150	NPK	-0.01295	SOL	-0.0007
DEL	-0.0194	PAM	-0.0288	TBS	-0.0077
DLV	0.0505	PEP	0.0014	TNT	-0.0019
ECO	-0.0017	POW	-0.0058	TRE	0.0808
ELH	-0.0163	PPC	-0.0027	UTR	0.0282
FIN	-0.0102				

GLS Transformed Regression			
R-squared	-0.0371	Mean dependent var	-0.0369
Adjusted R-squared	-0.0852	S.D. dependent var	0.1837
S.E. of regression	0.1914	Sum squared resid	7.1040
Durbin-Watson stat	1.6914		
Unweighted Statistics including Random Effects			
R-squared	-0.1010	Mean dependent var	-0.0368
Adjusted R-squared	-0.1521	S.D. dependent var	0.1837
S.E. of regression	0.1972	Sum squared resid	7.5420
Durbin-Watson stat	1.5932	Total Observations	204

Table B.4: **Balanced Panel Random Effects Regression Output for the Absolute Percentage Error** – A balanced panel is constructed with all firms with a full set of explanatory variables for the period 1994–1999. Firms reporting less than 20 cents in earnings are eliminated from the sample, as are two firms with HVOL values more than four standard deviations from the mean. The result is a 34 firm panel. The following model is then estimated: $ABSPER_{it} = \sum_{k=1}^{K} \beta_k \mathbf{X}_{kit} + \epsilon_{it}, \quad \epsilon_{it} = \alpha_i + \eta_{it}$

Variable	Coefficient	p Value
Intercept	-0.0723	0.6312
EY	0.2965	0.4533
LNMC	-0.0050	0.5663
DY	-0.2825	0.6586
HGRO	0.0143	0.8845
HVOL	0.0114	0.7835
ABSPER_L	0.3251	**0.0019**
INF	0.7336	0.2105
PRIME	0.6614	0.3066
BETA	-0.0179	0.5745

Firm	Intercept	Firm	Intercept	Firm	Intercept
AFE	0.0032	FOS	-0.0024	RCH	0.0003
AFX	0.0003	HVL	0.0012	RLO	0.0059
AGL	0.0034	IMP	-0.0038	RLY	0.0140
ALT	0.0150	JDG	0.0010	RMT	0.0066
AMS	-0.0265	KER	0.0094	SAP	0.0028
ATN	0.0018	LTA	0.0142	SIS	0.0012
DBR	-0.0036	NPK	0.0071	SOL	0.0084
DEL	0.0076	PAM	0.0037	TBS	0.0073
DLV	-0.0401	PEP	-0.0134	TNT	0.0099
ECO	0.0041	POW	0.0026	TRE	-0.0507
ELH	0.0060	PPC	0.0046	UTR	-0.0159
FIN	0.0148				

GLS Transformed Regression			
R-squared	0.0104	Mean dependent var	0.0991
Adjusted R-squared	-0.0355	S.D. dependent var	0.1589
S.E. of regression	0.1616	Sum squared resid	5.0689
Durbin-Watson stat	1.7518		

Unweighted Statistics including Random Effects			
R-squared	-0.0394	Mean dependent var	0.0991
Adjusted R-squared	-0.0877	S.D. dependent var	0.1589
S.E. of regression	0.1657	Sum squared resid	5.3244
Durbin-Watson stat	1.6677	Total Observations	204

Appendix C

Appendix to Chapter 6b – Out of sample testing

Out of sample testing for the balanced panel regression for the signed percentage error is attempted in two time periods. There are a number of problems with the data used for the testing, however the results may still be of interest.

The fixed effects panel results suggest a one factor model where the current signed percentage error is explained by the lagged error:

$$PERA_{i,t} = \alpha_i + \beta_i PERA_{i,t-12} \tag{C.1}$$

By estimating the model using only observations where the firms have positive earnings and positive earnings forecasts, the signed percentage error is given by:

$$PERA_{it} = \frac{A_{it} - F_{i,t-1}}{A_{it}} \tag{C.2}$$

PERA differs from the previously defined SURPE in the omission of the absolute value from the denominator.

These relationships can be rewritten as:

$$\frac{A_{it} - F_{i,t-1}}{A_{it}} = \alpha_i + \beta_{i,t-12} \tag{C.3}$$

This can be solved for the earnings at time t:

$$A_{i,t} = \frac{F_{i,t-1}}{1 - \alpha_i - \beta_{i,t-12}} \tag{C.4}$$

This model can be used to try and improve the analysts' forecasts using the forecast error from the previous period. If a less biased forecast can be produced without increasing the variance of the errors, this will represent a significant violation of the rational expectations hypothesis.

The first period of out of sample testing is from 1991–1993. While data are not available for the full set of explanatory variables for the period 1990–1993, the lagged error term is available for a number of the firms in the panel sample for this period. The potential weakness of this methodology is that the South African political environment underwent significant change during the early 1990s. It may therefore be inappropriate to estimate a model in the last part of the 1990s and test it in the first part. The second out of sample testing is based on estimating the model from 1994–1998 and testing the model for the full sample using the 1999 data.

The choice of the data used leaves the results open to criticism. For the first period the data needed to improve the forecasts were not available when the forecasts were made, and analysts had no way of incorporating these data into their forecasts. For the second period the 1999 data have already been used in the analysis that found the lagged error to be a significant predictor of the current error. It is likely that the forecast errors will improve the 1999 forecasts. Despite these criticisms, the results are potentially of interest.

C.0.1 Period 1, 1991–1993

The out of sample testing for this period is based on the fixed effects model re-estimated for the 1994–1999 period with PERA as the only independent variable. Modified earnings forecasts were then produced for the 1991–1993 period using

Table C.1: **Out of Sample Testing, 1991-1993** – A one factor model is estimated for the signed forecast error at time t, with the lagged error term as the explanatory variable. The sample used is the balanced panel sample. Solving for the earnings at time t gives: $A_{it} = \frac{F_{i,t-1}}{1-\alpha_i-\beta(PERA_{i,t-12})}$. Analysts' forecasts are then re-estimated for the period 1991-1993 using (6.11). The new forecasts are compared to the analysts' forecasts. Four firms are only able to provide two annual observations for the 1991-1993 period, the remaining 30 provide three observations each. The Wilcoxon Signed Rank Test T is a test of the null hypothesis that the modified forecasts' absolute percentage errors are less than or equal to the analysts' errors. The alternative hypothesis is that the modified model produces errors greater than the analysts.

	Model	**Analysts**
Mean	1.30%	-1.56%
Standard Deviation	13.02%	11.94%
Median	0.59%	-0.02%
Absolute Percentage Error	8.62%	7.34%
N	98	98
t-statistic		2.973

C.4. Four firms only had lagged errors available for two years, while the rest had lagged errors for three years. The summary statistics for the comparison of the analysts and the modified forecasts are presented in Table C.1. For the sample the analysts' mean signed error is -1.56% with a standard deviation of 11.94%. The forecast generated by the model had a mean percentage error of 1.3% with a standard deviation of 13.02%. The absolute percentage errors favour the analysts, with a mean error of 7.34% against a mean error of 8.62% for the modified forecast. The observation that the adjustment only improves 29 of the 99 forecasts is confirmed: the null hypothesis that the analysts' median error is equal to or greater than the adjusted forecast can be rejected at the 1% significance level (Wilcoxon signed rank test, T value 2.973).

Table C.2: Out of Sample Testing, 1999 – A one factor model is estimated for the signed forecast error at time t, with the lagged error term as the explanatory variable. The sample used is the balanced panel sample. Solving for the earnings at time t gives: $A_{it} = \frac{F_{i,t-1}}{1-\alpha_i-\beta(PERA_{i,t-12})}$. Analysts' forecasts are then re-estimated for 1999 using (6.11). The new forecasts are compared to the analysts' forecasts. The Wilcoxon Signed Rank Test T is a test of the null hypothesis that the modified forecasts' absolute percentage errors are less than or equal to the analysts' errors. The alternative hypothesis is that the modified model produces errors greater than the analysts.

	Model	**Analysts**
Mean	-0.17%	-4.89%
Standard Deviation	27.99%	32.73%
Median	4.84%	2.99%
Absolute Percentage Error	16.60%	16.55%
N	34	34
t-statistic		1.274

C.0.2 Period 2, 1999

The out of sample testing for this period is based on the fixed effects model re-estimated for the 1994–1998 period with PERA as the only independent variable. Modified earnings forecasts were then produced for the 1999 using C.4. The sample is identical to the balanced panel sample, 34 firms with one observation per firm. The results are presented in Table C.2

Again the model produces less biased results. The dispersion of analysts forecasts is however greater than that of the modified forecast. The analysts produce a smaller absolute percentage error; however the difference is very small. The Wilcoxon signed rank test T value is 1.274, indicating that there is no significant difference between the analysts and the modified forecast. Twenty two forecasts are made worse, while twelve improve.

For the four firms that had either two or three years of forecasts improved by the model in the 1991–1993 period, the analysts produced more accurate forecasts in 1999 than the model.

Conclusions

Despite the shortcomings of the methodology, the results do not provide evidence of violations of the rational expectations hypothesis. The forecasts cannot be improved except by increasing the variance of the forecast errors.

Appendix D

Appendix to Chapter 7 – Ordinary least squares and random effects regression output for the unbalanced panel and fixed effects regression omitting HVOL and ABSPER_L

Table D.1: **Ordinary Least Squares Regression Output for the Signed Percentage Error** – An unbalanced panel is constructed with all firms with a full set of explanatory variables for a minimum of four years during the 1991–1999 period. Firms reporting less than 20 cents in earnings are eliminated from the sample, as are two firms with HVOL values more than four standard deviations from the mean. The result is a 34 firm panel. The following model is then estimated: $PERA_{it} = \alpha + \sum_{k=1}^{K} \beta_k \mathbf{X}_{kit} + u_{it}$.

Variable		**Coefficient**	**p Value**
Intercept		-0.0930	0.5437
EY		-0.1604	0.7015
LNMC		0.0120	0.268
DY		0.5696	0.4394
HGRO		-0.0407	0.6161
HVOL		0.0130	0.7340
PERA_L		0.0273	0.7675
INF		0.3415	0.4745
PRIME		-0.3941	0.5539
BETA		0.0061	0.8618
R-squared	0.0134	Mean dependent var	-0.0352
Adjusted R-squared	-0.0199	S.D. dependent var	0.1962
S.E. of regression	0.1981	Sum squared resid	10.4769
F-statistic	0.4021	Durbin-Watson stat	1.6407
Prob(F-statistic)	0.9334	Observations	277

Table D.2: **Ordinary Least Squares Regression Output for the Absolute Percentage Error**– An unbalanced panel is constructed with all firms with a full set of explanatory variables for a minimum of four years during the 1991-1999 period. Firms reporting less than 20 cents in earnings are eliminated from the sample, as are two firms with HVOL values more than four standard deviations from the mean. The result is a 43 firm panel. The following model is then estimated: $ABSPER_{it} = \alpha + \sum_{k=1}^{K} \beta_k \mathbf{X}_{kit} + u_{it}$.

Variable		**Coefficient**	**p Value**
Intercept		0.04614	0.7318
EY		0.2121	0.5654
LNMC		-0.0058	0.5376
DY		-0.2746	0.6704
HGRO		0.0253	0.7228
HVOL		0.0111	0.7451
ABSPER_L		0.2553	**0.0138**
INF		-0.2973	0.4783
PRIME		0.4921	0.3974
BETA		-0.0071	0.8173
R-squared	0.0376	Mean dependent var	0.0977
Adjusted R-squared	0.0051	S.D. dependent var	0.1736
S.E. of regression	0.1732	Sum squared resid	8.0053
F-statistic	1.1584	Durbin-Watson stat	1.5921
Prob(F-statistic)	0.3221	Observations	277

Table D.3: **Unbalanced Panel Random Effects Regression Output for the Signed Percentage Error** – An unbalanced panel is constructed with all firms with a full set of explanatory variables for a minimum of four years during the 1991-1999 period. Firms reporting less than 20 cents in earnings are eliminated from the sample, as are two firms with HVOL values more than four standard deviations from the mean. The result is a 43 firm panel. The following model is then estimated: $PERA_{it} = \sum_{k=1}^{K} \beta_k \mathbf{X}_{kit} + \epsilon_{it}, \quad \epsilon_{it} = \alpha_i + \eta_{it}$.

Variable	Coefficient	p Value
Intercept	-0.1011	0.5166
EY	-0.1588	0.7077
LNMC	0.0136	0.2509
DY	0.5196	0.4924
HGRO	-0.0405	0.6201
HVOL	0.0224	0.5752
PERA_L	-0.0138	0.8817
INF	0.3093	0.5135
PRIME	-0.3822	0.5621
BETA	-0.0018	0.9618

Firm	Intercept	Firm	Intercept	Firm	Intercept
ABI	0.0177	FOS	-0.0153	PPC	0.0038
AFE	0.0019	HDC	0.0088	RCH	0.0159
AFX	-0.0122	HVL	0.0088	RLO	0.0076
AGL	0.0014	IMP	0.0186	RLY	0.0166
ALT	0.0095	JDG	-0.0031	RMT	0.0064
AMS	-0.0165	KER	0.0054	SAP	0.0074
ATN	0.0025	LTA	0.0152	SHP	-0.0707
CAS	0.0109	MDC	-0.0072	SIS	-0.0078
DBR	-0.0140	MUR	-0.0052	SOL	-0.0018
DDT	0.0175	NPK	0.01074	TBS	0.0064
DEL	0.0152	PAM	0.0222	TNT	0.0098
DLV	-0.0420	PEP	-0.0032	TRE	-0.0692
ECO	0.0040	PIK	0.0119	UTR	-0.0277
ELH	0.0151	POW	0.0048	WLO	0.0100
FIN	0.0101				

GLS Transformed Regression			
R-squared	0.0562	Mean dependent var	-0.0352
Adjusted R-squared	0.0244	S.D. dependent var	0.1962
S.E. of regression	0.1937	Sum squared resid	10.0223
Durbin-Watson stat	1.6651		

Unweighted Statistics including Random Effects			
R-squared	0.0910	Mean dependent var	-0.0352
Adjusted R-squared	0.0604	S.D. dependent var	0.1962
S.E. of regression	0.1901	Sum squared resid	9.6523
Durbin-Watson stat	1.7289	Observations	277

Table D.4: **Unbalanced Panel Random Effects Regression Output for the Absolute Percentage Error**– An unbalanced panel is constructed with all firms with a full set of explanatory variables for a minimum of four years during the 1991-1999 period. Firms reporting less than 20 cents in earnings are eliminated from the sample, as are two firms with HVOL values more than four standard deviations from the mean. The result is a 43 firm panel. The following model is then estimated: $ABSPER_{it} = \sum_{k=1}^{K} \beta_k \mathbf{X}_{kit} + \epsilon_{it}, \quad \epsilon_{it} = \alpha_i + \eta_{it}$.

Variable	Coefficient	p Value
Intercept	0.0955	0.4925
EY	0.2626	0.4821
LNMC	-0.0103	0.3545
DY	-0.3390	0.6152
HGRO	0.0235	0.742
HVOL	-0.0170	0.6403
ABSPER_L	0.1306	0.2161
INF	-0.2583	0.5281
PRIME	0.4622	0.4152
BETA	0.0003	0.9930

Firm	Intercept	Firm	Intercept	Firm	Intercept
ABI	-0.0167	FOS	0.0061	PPC	-0.0171
AFE	-0.0015	HDC	-0.0366	RCH	0.0032
AFX	0.0061	HVL	0.0028	RLO	-0.0232
AGL	-0.0067	IMP	0.0057	RLY	-0.0294
ALT	-0.0331	JDG	-0.0051	RMT	-0.0165
AMS	0.04981	KER	-0.0225	SAP	-0.0029
ATN	-0.0045	LTA	-0.0366	SHP	0.1221
CAS	-0.0283	MDC	-0.0115	SIS	-0.0080
DBR	0.0096	MUR	0.0407	SOL	-0.0139
DDT	-0.0120	NPK	-0.0220	TBS	-0.0198
DEL	-0.0195	PAM	0.0026	TNT	-0.0095
DLV	0.0987	PEP	0.0145	TRE	0.1042
ECO	-0.0117	PIK	-0.0259	UTR	0.0430
ELH	-0.0195	POW	-0.0079	WLO	-0.0040
FIN	-0.0324				

GLS Transformed Regression			
R-squared	0.1273	Mean dependent var	0.0977
Adjusted R-squared	0.0979	S.D. dependent var	0.1736
S.E. of regression	0.1649	Sum squared resid	7.2587
Durbin-Watson stat	1.6098		

Unweighted Statistics including Random Effects			
R-squared	0.1901	Mean dependent var	0.0978
Adjusted R-squared	0.1628	S.D. dependent var	0.1736
S.E. of regression	0.1588	Sum squared resid	6.7368
Durbin-Watson stat	1.7345	Observations	277

Table D.5: **Fixed Effects Regression Output for the Absolute Percentage Error Omitting HVOL** – An unbalanced panel is constructed with all firms with a full set of explanatory variables for a minimum of four years during the 1991–1999 period. Firms reporting less than 20 cents in earnings are eliminated from the sample, as are two firms with HVOL values more than four standard deviations from the mean. The result is a 43 firm panel. The following model is then estimated: $ABSPER_{it} = \alpha_i + \sum_{k=1}^{K} \beta_k \mathbf{X}_{kit} + u_{it}$. HVOL is omitted from the explanatory variables.

Variable	Coefficient	p Value
EY	-0.1990	0.6451
LNMC	-0.0781	**0.0214**
DY	-0.2747	0.7453
HGRO	-0.0150	0.8421
ABSPER_L	-0.2692	**0.0202**
INF	-0.7882	0.0813
PRIME	0.7577	0.2236
BETA	0.0365	0.4821

Firm	Intercept	Firm	Intercept	Firm	Intercept
ABI	0.6086	FOS	0.7003	PPC	0.5885
AFE	0.6781	HDC	0.4703	RCH	0.8268
AFX	0.6984	HVL	0.6222	RLO	0.5439
AGL	0.7860	IMP	0.6904	RLY	0.4285
ALT	0.4914	JDG	0.6109	RMT	0.7194
AMS	0.8971	KER	0.5880	SAP	0.7093
ATN	0.5832	LTA	0.4225	SHP	0.9832
CAS	0.5386	MDC	0.6282	SIS	0.6527
DBR	0.8530	MUR	0.8111	SOL	0.7750
DDT	0.7063	NPK	0.6303	TBS	0.6576
DEL	0.4907	PAM	0.6914	TNT	0.6563
DLV	0.9586	PEP	0.7102	TRE	0.9686
ECO	0.6470	PIK	0.4855	UTR	0.7797
ELH	0.5489	POW	0.5844	WLO	0.6687
FIN	0.4339				

R-squared	0.3362	Mean dependent var	0.0977
Adjusted R-squared	0.1893	S.D. dependent var	0.1736
S.E. of regression	0.1563	Sum squared resid	5.5218
F-statistic	16.3484	Durbin-Watson stat	1.7613
Prob(F-statistic)	0	Total Observations	277

Table D.6: **Fixed Effects Regression Output for the Absolute Percentage Error Omitting ABSPER_L** – An unbalanced panel is constructed with all firms with a full set of explanatory variables for a minimum of four years during the 1991–1999 period. Firms reporting less than 20 cents in earnings are eliminated from the sample, as are two firms with HVOL values more than four standard deviations from the mean. The result is a 43 firm panel. The following model is then estimated: $ABSPER_{it} = \alpha_i + \sum_{k=1}^{K} \beta_k \mathbf{X}_{kit} + u_{it}$. ABSPER_L is omitted from the explanatory variables.

Variable	**Coefficient**	**p Value**
EY	-0.0020	0.9961
LNMC	-0.1069	**0.0021**
DY	-0.9549	0.2607
HGRO	0.0377	0.6092
HVOL	-0.1711	**0.0003**
INF	-0.5051	0.2513
PRIME	0.9766	0.1133
BETA	0.0355	0.4863

Firm	**Intercept**	**Firm**	**Intercept**	**Firm**	**Intercept**
ABI	0.7835	FOS	0.8685	PPC	0.7775
AFE	0.8610	HDC	0.5905	RCH	1.0594
AFX	0.8473	HVL	0.8775	RLO	0.7241
AGL	1.0398	IMP	0.9107	RLY	0.6800
ALT	0.6514	JDG	0.7871	RMT	0.9452
AMS	1.1051	KER	0.7799	SAP	1.0272
ATN	0.7150	LTA	0.5613	SHP	1.2750
CAS	0.6899	MDC	0.7192	SIS	0.8415
DBR	1.1274	MUR	1.0851	SOL	0.9969
DDT	0.8780	NPK	0.8362	TBS	0.8500
DEL	0.6279	PAM	0.9322	TNT	0.8499
DLV	1.1136	PEP	0.9333	TRE	1.1279
ECO	0.8359	PIK	0.6441	UTR	0.8815
ELH	0.7019	POW	0.7283	WLO	0.8630
FIN	0.6663				

R-squared	0.3591	Mean dependent var	0.0977
Adjusted R-squared	0.2174	S.D. dependent var	0.1736
S.E. of regression	0.1536	Sum squared resid	5.33064
F-statistic	18.0923	Durbin-Watson stat	1.9368
Prob(F-statistic)	0	Total Observations	277

Bibliography

[1] ALBRECHT W.S., JOHNSON O., LOOKABILL L.L. & WATSON D.J.H. (1977). A Comparison of Corporate and Security Analysts' Forecasts of Earnings: A Comment. *The Accounting Review* **52**(3),736–740.

[2] ALBRECHT W.S., LOOKBILL L.L. & MCKEOWN J.C. (1977). The time series properties of earnings. *Journal of Accounting Research* Autumn, 226–244.

[3] ALLEN A., CHO J.Y. AND JUNG K. (1997). Earnings forecast errors: Comparative evidence from the Pacific-Basin capital markets. *Pacific-Basin Finance Journal* **5**, 115–129.

[4] AMIR E., LEV B., & SOUGIANNIS T. (1999) What value analysts? Working paper.

[5] ATTFIELD C.L.F. DEMERY D. & DUCK N.W. (1985). Rational Expectations in Macroeconomics. Basil Blackwell, Oxford.

[6] BALL R.(1989). Ball and Brown[1968]. *Journal of Accounting Research* **27**(3), 202–217.

[7] BALL R. & BROWN P. (1968). An Empirical Evaluation of Accounting Numbers. *Journal of Accounting Research* **6**(2), 159–178.

[8] BALL R. AND WATTS R. (1972). Some Time Series Properties of Accounting Income. *Journal of Finance* June, 663–681.

[9] BIRD R., MCELWEE B. & MCKINNON J. (2000). A Global Perspective of Analysts' Earnings Forecasts. *Journal of Investing* **9**(4), 76–83.

[10] Black F. (1993). Beta and Return. *Journal of Portfolio Management* **20**, 8–18.

[11] Brooks L.D. & Buckmaster D.A. (1976). Further Evidence on the Time Series Properties of Accounting Income. *The Journal of Finance* **31**(5), 1359–1373.

[12] Brown P. (1989). Invited Remarks: Ball and Brown [1968]. *Journal of Accounting Research* (Supplement) **27**, 202–217.

[13] Brown P., Foster G. & Noreen E. (1985). Security Analyst Multi-Year Earnings Forecasts and the Capital Markets. *American Accounting Association, Fl.*.

[14] Brown L.D. (1996). Analyst Forecasting Errors and Their Implications for Security Analysis: An Alternative Perspective. *Financial Analysts Journal* Jan/Feb, 40–47.

[15] Brown L.D. (1997*a*). Earnings Surprise Research: Synthesis and Perspectives. *Financial Analysts Journal* March, April, 13–19.

[16] Brown L.D. (1997*b*). Analyst Forecasting Errors: Additional Evidence. *Financial Analysts Journal* November/December, 81–88.

[17] Brown L.D. and Rozeff M.S. (1978). The Superiority of Analyst Forecasts as Measures of Expectations: Evidence from Earnings. *The Journal of Finance* **33**(1), 1–16.

[18] Brown L.D., Richardson G.D. & Schwager S.J. (1987). An Information Interpretation of Financial Analyst Superiority in Forecasting Earnings. *Journal of Accounting Research* **25**(1), 49–67.

[19] Campbell J.Y. (2000). Asset Pricing at the Millenium. *The Journal of Finance* **55**(4) Papers and Proceedings of the American Finance Association, Boston, Massachusetts, January 7–9,2000, 1515–1567.

[20] CAPSTAFF J., PAUDYAL K. & REES W. (2001). A Comparative Analysis of Earnings Forecasts in Europe. *Journal of Business Finance and Accounting* **28**(5), 531–562.

[21] CAPAUL C., ROWLEY I. & SHARPE W.F. (1993). International value and growth stock returns. *Financial Analysts Journal* **49**, 27–36.

[22] CHAN L.K.C., HAMAO Y. AND LAKONISHOK J. (1991). Fundamentals and stock returns in Japan. *Journal of Finance* **46** (5), 1739–1789.

[23] CHANG J.J., KHANNA T. AND PALEPU K. (2000). Analyst Activity Around the World. Working paper.

[24] CONOVER W.J. (1980). Practical Nonparametric Statistics. John Wiley and Sons.

[25] CONROY R.M. & HARRIS R.S. (1995). Analysts' earnings forecasts in Japan: Accuracy and sell-side optimism. *Pacific-Basin Finance Journal* **3**,393–408.

[26] COPELAND T. & WESTON F. (1988). Financial Policy and Corporate Theory, 3rd Edition. Addison Wesley.

[27] CRAGG J.G. & MALKIEL B.G. (1968). The Consensus and Accuracy of Some Predictions of the Growth of Corporate Earnings. *The Journal of Finance* **23**(1),67–84.

[28] CRICHFIELD T., DYCKMAN T. & LAKONISHOK J. (1978). An Evaluation of Security Analysts' Forecasts. *The Accounting Review* **53**(3), 651–668.

[29] DANIEL K. & TITMAN S. (1997). Evidence on Characteristics of Cross Sectional Variation in Stock Returns . *Fournal of Finance* **52** (1), 1–33.

[30] DAS S., LEVINE C.B. & SIVARAMAKRISHNAN. (1998). Earnings Predictability and Bias in Analysts' Earnings Forecasts. *The Accounting Review* **73** (2), 277–294.

[31] DAVIS, FAMA E. & FRENCH K. (2000). Characteristics, Covariances, and Average Returns: 1929 to 1997. *Journal of Finance* **55** (1), 389–406.

[32] De Bondt W.F.M. & Thaler R.H. (1987). Further Evidence on Investor Overreaction and Stock Market Seasonality. *Journal of Finance* **42**(2), 557–581.

[33] De Bondt W.F.M. & Thaler R.H. (1990). Do Security Analysts Overreact? *The American Economic Review* **80** (2), 52–57.

[34] Deschamps B. & Mehta D.R. (1980). Predictive Ability and Descriptive Validity of Earnings Forecasting Models. *The Journal of Finance* **35**(4), 933–949.

[35] Dreman D.N. & Berry M.A. (1995). Analyst Forecasting Errors and Their Implications for Security Analysis. *Financial Analysts Journal* May/June, 30–41.

[36] Easley D. & Jarrow R.A. (1983). Consensus Beliefs Equilibrium and Market Efficiency. *The Journal of Finance* **38**(3), 903–911.

[37] Elton E.J. & Gruber M.J. (1972). Earnings Estimates and the Accuracy of Expectational Data. *Management Science* **18** (80), B409–B424.

[38] Elton E.J., Gruber M.J. & Gultekin M.N. (1984). Professional Expectations: Accuracy and Diagnosis of Errors. *Journal of Financial and Quantitative Analysis* **19**(4),351–363.

[39] Fama E.F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance* **25**(2), 383–417.

[40] Fama E.F. & French K. (1992). The Cross Section of Expected Stock Returns. *Journal of Finance* **47**(2), 427–465.

[41] Fama E.F. & French K. (1993). Common risk Factors in the Returns to Stocks and Bonds. *Journal of Financial Economics* **33**, 3–56.

[42] Fama E.F. and French K. (1995). Size and Book-to-Market factors in Earnings and Returns. *Journal of Finance* **50** (1), 131–155.

[43] Fama E.F. and French K. (1998). Value Versus Growth: the International Evidence. *The Journal of Finance* **53** (6), 1975–1999.

[44] Fama E.F. & Miller M.H. (1972). The Theory of Finance. Holt, Rinehart and Winston.

[45] Foster G. (1986). Financial Statement Analysis. Prentice Hall.

[46] Foster G., Olsen C. & Shevlin T. (1984). Earnings Releases, Anomalies, and the Behavior of Security Returns. *Accounting Review* **59**(4), 574–603.

[47] Freeman R.N. and Tse S. (1989). The Multiperiod Information Content of Accounting Earnings Confirmations and Contradictions of Previous Earnings Reports. *Journal of Accounting Research*(Supplement) **27**, 49–84.

[48] Fried D. and Givoly D. (1982). Financial Analysts' Forecasts of Earnings. *Journal of Accounting and Economics* **4**, 85–107.

[49] Haugen R. (1995). The New Finance: The Case against Efficient Markets. Prentice Hall.

[50] Higgins H.N. (1998). Analyst Forecasting Performance in Seven Countries. *Financial Analysts Journal* May/June, 58–62.

[51] Higgins H.N. (2002). Analysts' forecasts of Japanese firms' earnings: additional evidence. *The International Journal of Accounting* **37**, 371–394.

[52] Hsiao C. (2003) Analysis of Panel Data. Cambridge University Press.

[53] Johnston J. and DiNardo J. (1997). Econometric Methods, 4th Edition. McGraw Hill.

[54] Kantor B. (1979). Rational Expectations and Economic Thought. *Journal of Economic Literature* **17**(4), 1422–1441.

[55] Keane M.P. & Runkle D.E. (1998). Are Financial Analysts' Forecasts of Corporate Profits Rational? *The Journal of Political Economy* **106**(4), 768–805.

[56] Kim D. & Kim M. (2003). A Multifactor Explanation of Post-Earnings Announcement Drift. *Journal of Financial and Quantitative Analysis* **38**(2), 383–398.

[57] KROSS W., RO B. & SCHROEDER D. (1990). Earnings Expectations: The Analysts' Information Advantage. *The Accounting Review* **65**(2), 461–476.

[58] LAKONISHOK J., SHLEIFER A. AND VISHNY R.W. (1994).Contrarian Investment, Extrapolation and Risk. *Journal of Finance* **49**(5), 1541–1578.

[59] LOVELL M. C. (1986). Tests of the Rational Expectations Hypothesis. *The American Economic Review* **76**(1), 110–124.

[60] MCDONALD C.L. (1973). An Empirical Examination of the Reliability of Published Predictions of Future Earnings. *The Accounting Review* **48**(3), 502–510.

[61] MACKINLAY A.C. (1995). Multifactor models do not explain deviations from the CAPM. *Journal of Financial Economics* **38**, 3–28.

[62] MILLER M.H. AND MODIGLIANI F. (1961). Dividend Policy, Growth and the Valuation of Shares. *The Journal of Business* **34**(4), 411–433.

[63] MODIGLIANI F. AND MILLER M.H. (1958). The Cost of Capital, Corporation Finance and the Theory of Investment. *The American Economic Review* **48**(3), 261–297.

[64] MUTH J.F. (1960). Optimal Properties of Exponentially Weighted Forecasts. *Journal of the American Statistical Association* **55**(290), 299–306.

[65] MUTH J.F. (1961). Rational Expectations and the Theory of Price Movements. *Econometrica* **29**(3), 315–335.

[66] NIEDERHOFFER V. AND REGAN P. (1972). Earnings Changes, Analysts Forecasts, and Stock Prices. *Financial Analysts Journal* May/June, 65–71.

[67] PARKESH M., DHALIWAL D.S. & SALATKA W.K. (1995). How Certain Firm-Specific Characteristics Affect the Accuracy and Dispersion of Analysts' Forecasts. *Journal of Business Research* **34**, 161–169.

[68] RICHARDS R.M. (1976) Analysts Performance and the Accuracy of Corporate Earnings Forecasts. *The Journal of Business* **49**(3), 350–357.

[69] RICHARDS R.M., BENJAMIN J.J. & STRAWSER RH. (1977). An Examination of the Accuracy of Earnings Forecasts. *Financial Management* Fall, 78–86.

[70] VAN RENSBURG P. AND ROBERTSON M. (2003). Style Characteristics and the Cross-section of JSE Returns. *Investment Analysts Journal* **57**, 7–16.

[71] VAN RENSBURG P. (2001). A Decomposition of Style Based Risk on the JSE. *Investment Analysts Journal* **54**.