Stockholm School of Economics
Department of Accounting
Master Thesis in Accounting and Financial Management

Are Security Analyst Forecasts Bulletproof?

A Study of Security Analysts' Accuracy in Forecasting Quarterly Earnings

Carlsson Sven (19856) Hillerborg Anders (19901)

Abstract

The purpose of this Master Thesis is to evaluate the contribution of Security Analysts' earnings forecasts, by comparing their forecasting accuracy with that of Time-Series Models, and to test if and how their forecasting accuracy varies with different circumstances. We have studied Security Analysts' forecasts on quarterly earnings of Swedish companies between 2001 and 2007. We have measured the forecasting accuracy in two ways; (1) using Percentage Error Analysis, which has been the most commonly used evaluation tool in previous research, and (2) Regression Line Analysis, which is used in order to reduce the measurement bias caused by the size of earnings when using the Percentage Error approach. We find that Security Analysts are more accurate than Time-Series Models when forecasting quarterly earnings, although they tend to underestimate their earnings forecasts. The forecasting accuracy of security analysts varies with different circumstances. We conclude that the accuracy varies between different sectors, and that the most accurate forecasts are for companies in the Healthcare sector. For the remaining circumstances investigated in this thesis, namely the size of the company, the forecasted quarter and the market trend, we do not find any significant differences in Security Analysts' forecasting accuracy.

Keywords: Security Analyst Forecast, Univariate Time-Series Models, Earnings Forecast, Forecasting Accuracy, Earnings Forecasting Accuracy

Tutor:
Skogsvik Stina

Presentation Date
22nd May 2008
Room 542 10-12

Acknowledgements

There are several people who have shown a great deal of support and without whose aid this thesis would not have been finalized. We would like to start off by showing gratitude to our tutor Stina Skogsvik who has acted as a beacon guiding us back on track when so needed. We also want to thank Per-Olov Edlund from the Department of Economics Statistics and Decision Support for his tireless support and input. We would also like to take this opportunity to show our gratitude to our respective parents, who have supported us in all our endeavors through our lives.

Index

1-Introduction

1.1 Background

Much research has been done in the field of different valuation models, such as Free Cash Flow, Adjusted Present Value, Value Added and Residual Income Valuation. These models deal with how to value a company based on assumptions of future performance, measured as for example free cash flow or earnings measures. They do not address one of the most crucial components of valuation: how the forecasts that will be inserted into the valuation models are produced.

Previous research has shown that earnings forecasts are vital both for company valuation and stock price changes.[1] Earnings forecasts can be either short term, as for example quarterly or annual earnings, or long term, as for example five-year earnings growth rates. [2] For the purpose of this thesis, we will focus on the shorter perspective, in form of quarterly earnings forecasts. There are several different research areas on earnings forecasting including but not limited to stock market reaction to earnings forecasts and forecast errors, the attributes of managements' earnings forecasts and the attributes of Security Analysts' earnings forecasts. We will focus on the latter: the attributes of Security Analysts' earnings forecasts.

Users of earnings forecasts normally rely on the forecasts produced by Security Analysts. There is however a risk that too much trust is being put on these earnings forecast without the proper knowledge regarding the accuracy and reliability of the forecasts and how it may vary between different circumstances. Furthermore, the existence of simple extrapolating Time-Series Models that operate as a low cost alternative to the more expensive Security Analysts Earnings forecasts[3] and which have become widely used indicates that there is a need to explore the contribution made by Security Analysts.

To the best of our knowledge, no similar evaluation of Security Analysts' earnings forecasting has been carried out on Swedish companies and Security Analysts covering Swedish companies based on quarterly data before. Therefore, we believe that we will be able to

[1] Brown & Rozeff (1978) pp.1-16
[2] Foster (1986) p.262
[3] Foster (1986) p.278

contribute valuable insight about the quality and characteristics of Security Analysts' earnings forecasts.

1.2 Purpose

In order to find out the contribution of Security Analysts' forecasts to investors, the purpose of this study is to (1a) compare the accuracy of earnings forecasts for Security Analysts in relation to Times-Series Models, and to (1b) explain and analyze if Security Analysts and Times-Series Models tend to overestimate or underestimate.

In the first part of our purpose (1a and 1b), we look at the entire sample, assuming that the performance by Security Analysts is the same in all circumstances. As this most likely is not the case, our second part of the purpose (2) is to test if and how Security Analysts' forecasting accuracy varies with different circumstances, namely with company size, sectors belonging, fiscal quarters and market trend.

The aim of this thesis is to draw conclusions that may be useful when dealing with earnings forecasts on a quarterly basis in Swedish companies. Hence we hope to improve the understanding of the performance and limitations of earnings forecasting. Since earnings forecasts are used by a wide variety of professionals, this study has a practical as well as a theoretical application. The main application lie in earnings forecasts for corporate valuation models. Therefore we primarily write this thesis for an investor. With an investor we primarily mean an institutional- or other professional investor.

1.3 Delimitations

There are countless circumstances that can be examined; however, we have decided in this thesis to focus on: Size, Sector, Quarter and Market Trend. The decision to focus on these four circumstances is based on previous research and factors we assume to be of interest, in respect to our purpose.

It is, however, not our intent to analyze the effects on stock market returns or valuation that superior forecasts could generate.

1.4 Theses Outline

We will begin by looking at our theoretical framework, followed by a presentation of our hypotheses. Next, we will explain our methodology and definitions. After that we will present our empirical data followed by our analysis. Finally, we will draw our conclusion.

2- Theoretical framework

2.1. The Value of Forecasting

The value of forecasting was explored as early as 1968 in a study conducted by Ball & Brown. The study has been revisited several times by other scholars and its findings have been found to be robust both over time and between countries. The study investigated the value of accounting numbers. Ball & Brown drew three conclusions from their study: 1) Earnings information given in the accounting effects investors in their valuation 2) The effect of any changes in expectations was limited by the time at which it was announced 3) With superior knowledge of future earnings abnormal profits may be earned. This study is valuable to our thesis because Ball & Brown showed the link between stock value and accounting numbers, including earnings. Furthermore it gave the insight that the value created by accounting numbers has a declining effect as the announcement date gets closer and that given perfect information abnormal profits can be made. This showed that by making superior earnings forecasts an investor can make abnormal profits. It is also critical to make them as early as possible. [4] Foster further defined that "a better quality forecast is one that leads to a better quality decision being made". He specified four different decision contexts where good forecasts are important: 1) for an investor to find miss-priced securities, 2) for an investor to achieve a good diversification and with a particular risk level, 3) for a credit lender to match the interest level against the risk of a default, and 4) for a company to value a part of the company that is to be divested.[5]

We conclude that accounting numbers are important for generating forecasts. We narrow down accounting numbers to only include earnings numbers. Accounting number are however not always reliable as management have the possibility to affect earnings (for example by earnings management. Another effect comes from earnings numbers sometimes are not being accounted for in the correct accounting period.

[4] Ball & Brown (1968) pp.159-178
[5] Foster (1986) p.266

As described by Watts accountants divide cost over a fiscal year into quarterly reports based on previous ratios. This may not be the actual cost incurred, hence in the later quarters this is rectified which often results in earnings not being finalized until the fourth quarter.[6]

This implies that the fourth quarter's earnings numbers are the most unreliable. Besides this accounting effect, which we will be examined in more detail later, we assume no other issues, such as earnings management, that will significantly impact our earnings numbers. In the next section, we will continue by looking at the different approaches to forecasting.

2.2. Forecasting Approaches

Foster describes four different approaches of forecasting, which can be sorted into a matrix depending on if they are mechanical or non-mechanical and if they are univariate or multivariate. A mechanical approach combines the inputs in a pre-determined way. A non-mechanical approach, on the other hand, has no specified link between the forecast and the previous outcomes (a Times-Series Model). One example could be to take judgmental factors into consideration. The models could further be split into univarite and multivariate approaches. A univarite approach only considers one factor, whereas a multivariate approach incorporates several variables. [7]

Table 2.1 Forecasting Approaches

	Univariate	**Multivariate**
Mechanical	Moving average models Box-Jenkins univariate models	Regression models Box-Jenkins transfer function models Econometreic models
Nonmechanical	Visual curve extrapolation	Security analyst approach

Source:Foster (1986)p.262

We will focus on two of these four approaches: (1) The non mechanical and multivariate approach, and (2) the mechanical and univariate approach. The first approach is represented by Security Analysts, while the latter is represented by Times-Series Models.

Foster further outlines that both Times-Series Models and Security Analysts have their advantages and disadvantages. Security Analysts may incorporate information from several sources and may adjust forecasts when new information becomes available. However, they

[6] Watts (1976) pp.81-85
[7] Foster (1986) p.262

also come at a high cost and may include influences that can lead to biased forecasts. Times-Series Models on the other hand are cheap and easy to update, do not make subjective forecasts, and may find and take advantage of systematic patterns in past data, while the drawbacks of the Times-Series Models are their inability to produce updated forecasts between the earnings releases and to inability produce forecast for new companies without previous time series available.[8] [9]

When comparing the two approaches, previous studies, including those by Brown, Griffin, Hagerman & Zmijewski and Brown & Rozeff, have found that Security Analysts produce more accurate earnings forecasts than Times-Series Models. The study by Brown, Griffin, Hagerman & Zmijewski further showed that Security Analysts' superiority was not dependent on (1) chronological sub periods, (2) forecast horizon, (3) forecast error definition and definition of outliers, (4) conditional quarter, or (5) the statistical test used, as both parametric- and non-parametric tests give the same result.[10] Brown & Rozeff further add that the pure existence of Security Analysts implies that Security Analysts' forecast must be superior to time series models.[11]

We have shown that both models have their advantages and disadvantages, but conclude that Security Analysts have shown higher forecasting accuracy. The following two sections will explain the specific characteristics of Security Analysts forecasting (2.3) and Times-Series Models forecasting (2.4).

2.3. Security Analysts Forecasting

There are two main reasons for why Security Analysts are superior to the Times-Series Models. Firstly, several studies have shown that Security Analysts' forecasting errors decreases as the announcement date approaches, i.e. suggesting that there is a timing advantage. [12] [13] [14] While a Times-Series Model can't produce new forecast between the reports, Security Analysts may incorporate new information made available between the interim reports into their forecasts. Secondly, Security Analysts has the ability to incorporate

[8] Foster (1986) p.278
[9] For a complete list see Table A1.1
[10] Brown et.al. (1984) pp.61-87
[11] Brown & Rozeff (1978) pp.1-16
[12] Foster (1986) p.275
[13] Chrichfield, Dryckman &Lobon (1978) pp.661-668
[14] Brown et.al. (1984) pp.61.87

not only more up to date earnings indications but also incorporate several other factors than previous earnings. [15] One reason for this is that univariate Times-Series Models neglect potential useful information that lies outside its own time series. [16]Example of such information could be macroeconomic forecasts, industry structure, and special events such as acquisitions and acquisitions.[17] A study by Collins & Hopwood showed that Security Analysts not only had lower means and standard deviations than the Times-Series Models, but also fewer and lower extreme outliers. This indicates that Security Analysts are better at predicting the effects from extraordinary circumstances. It also implies that they are better at incorporating known information which should result in an increase in forecast accuracy.[18] Finally, Brown & Rozeff state that the existence of Security Analysts means that their forecast must be superior to those of Times-Series Models.[19]

One study that captured both the timing advantage and the ability to incorporate several sources of information was a study by Elton, Gruber and Gultekin that have examined the sources of Security Analysts' errors. The study looked at how three components affect the error: 1) Economy component (in this case the earnings of all firms in the sample), 2) Industry component (total earnings of an industry), and 3) Firm Component (the inability to predict how a firm differs from its industry average). The first conclusion is that Security Analysts' total errors decrease as the reporting date gets closer. Secondly, when looking at the sources of error, this study reveals that the majority of the errors were from the firm specific component, followed by the industry component and the economy component actually generates the fewest errors.. When comparing forecasts in January and December for the annual report, the Economy components share of the error declines from 2.0% to 0.8% and the Industry Component decreases from 37.3% to 15.5%, while the Company Component increases from 60.7% to 83.7%. Considering the findings that the absolute size of the error also declines with time, this indicates that Security Analysts are able to eliminate the major part of the error from the economy and industry components significantly over time (and most likely also the error for the Company Component).[20]

[15] Brown et.al. (1984) pp.61-87
[16] Brown & Rozeff (1978) pp.1-16
[17] Foster (1986) p.279
[18] Collins & Hopwood (1980) pp.390-406
[19] Brown & Rozeff (1978) pp.1.16
[20] Elton,Gruber &Gultekin (1984) pp.351-363

We conclude that Security Analysts seem to have two main advantages over Times-Series Models; a timing advantage and the ability to incorporate several sources of information. Although being more accurate than Times-Series Models, there are differences in Security Analysts' forecast accuracy between different circumstances.

Bathke, Lorek & Willinger studied the correlation between the size of the firm and the forecasting error produced by Security Analysts. They concluded that Security Analysts accuracy in forecasting earnings was dependant on the size of the firm but also that Security Analysts were significantly more accurate when forecast large and mid size firms than small size firms. Although the study also concluded that it may not be necessary to conduct separate analysis based on firm size they also concluded that size is a factor to consider. [21]

Next we look at previous research on consensus forecasts. The basis for using Security Analysts' forecasts on an aggregated level instead of an individual is quite straight forward because it relies on the assumption that the individual Security Analyst at some point in time shows errors in their forecasts. The errors are however not correlated to other Security Analysts' errors thus when aggregated the individual forecast errors are canceled out by others errors.

Several previous studies have shown that consensus forecasts perform better than individual Security Analysts' forecasts. A study of forecasts of GDP, the implicit price deflator and consumer expenditure for durable goods, has shown that consensus is more accurate than most individuals in each survey. Over time, no individual managed to achieve any long-term superiority over consensus. [22] Similar results on earnings per share have also been demonstrated.[23] The question of what the minimum number of Security Analysts to form a consensus group has been visited by several researchers i.e. Ashton & Ashton. They showed that a group consisting of a fairly small number of Security Analysts was able to make large improvements to the individual SE's forecasts.[24] Conroy & Harris conclude that the main benefit from the consensus approach is that large improvements may be gained with limited costs. [25]

[21] Barhke, Lorek & Willinger (1984) pp.49-58
[22] Foster (1986) p.285
[23] Foster (1986) pp.285-286
[24] Ashton & Ashton (1986) pp.1499-1508
[25] Conroy & Harris (1987) pp.725-738

We conclude that consensus forecasts are better to use than individual forecasts. There are furthermore two methods used to calculate consensus: the mean and the median. Both methods are widely used and often both methods are reported. The methods are slightly different as the mean captures the deviations in size whereas the median captures the middle value. Both methods have their disadvantages. The mean could be heavily distorted by single extreme values, while median could on the other hand be skewed if all values are extreme.

2.4. Time-Series Models Forecasting

There are several different ways of creating time-series models. We will now present some theories and evidence on two of the most common ones: The Martingale model and the Sub Martingale model.

The Sub Martingale suggests that the most recent observation, adjusted for trend (δ) is the best extrapolative estimate for future earnings. A Martingale model is a special case of the Sub Martingale with no drift factor (δ = 0), meaning that the next outcome is assumed to be the same as the previous observation. The Sub Martingale model is a first order autoregression [26]:

$Y_t = Ø_{yt-1} + δ + μ_t$

Where

$Y_1, Y_2 \ldots Y_n$ are successive observations

Ø = 1 is the weight placed upon the most recent observation

δ is an drift factor with value ≥0

$μ_t$ is a random disturbance

2.4.1 Performance of Martingale and Sub Martingale Models

As previously mentioned Times-Series Models operate as a low cost alternative and are widely used. Security Analysts often also base their forecasts on an extrapolation of the company's historical performance.[27] As illustrated by Ball and Watts, annual earnings for US firms follow "a Sub Martingale or some very similar process". [28] When choosing between a Sub Martingale model with a drift factor or the pure Martingale model previous studies have come to different conclusions depending on the time series, sample and measure used. Foster and Finn & Whittred conclude that the Sub Martingale model is better than the pure

[26] Ruland (1972)pp.90-37
[27] McEnallly (1971) pp.687-706
[28] Ball & Watts (1972) pp.663-682

Martingale, as the drift incorporate factors such as growth caused by inflation[29] and as the impact created when companies reinvested earnings to create growth.[30]

Previous research has compared the performance of industries for the Martingale and Sub Martingale model was unable to draw any clear conclusions since the intent of the research was not to compare sectors but rather to examine the effectiveness of different time-series models.[31]

Finn & Whittred concluded that both the martingale model and all the Sub Martingale models consistently underestimated realized earnings changes for the period 1971-1978. [32]

[29] Finn & Whittred (1982) pp.169-173
[30] Foster (1977) pp.1-21
[31] Ruland (1972) pp.30-37
[32] Finn & Whittred (1972) p.171

3 Hypotheses

In order to answer the questions outlined in our purposes we have formulated 6 hypotheses, each which is followed by a brief motivation. The first two hypotheses relates to our first purpose of how Security Analysts perform in relation to Times-Series Models, and the last four hypotheses are used to answer our second purpose to determine if Security Analysts' forecasting accuracy varies with different circumstances.

H1) *Security Analysts provide more accurate forecasts than Time-Series Models*

Previous research i.e. Brown, Griffin, Hagerman & Zmijewski in 1984 and Brown & Rozeff in 1978, has indicated that Security Analysts make more accurate forecasts than Times-Series Models. Security Analysts have an advantage of time since they are able to make forecasts closer to the reporting date.[33] [34]Security Analysts also have access to more information and may incorporate several factors except just previous earnings information,[35] [36] and are also better at predicting extraordinary items.[37] Thus, we expect Security Analysts to be able to generate better forecasts than Times-Series Models.

H2) *The Time-Series Models tend to underestimate earnings while Security Analysts neither overestimate nor underestimate earnings*

Previous research by Finn & Whittred supports that Times-Series Models consistently underestimated earnings. When we combine the advantages of Security Analysts outlined in hypothesis 1 with the fact there is a very competitive situation between analysts which motivates them to be as accurate as possible, we have reason to assume that Security Analysts should neither overestimate nor underestimate earnings. [38]

[33] Brown et.al.(1987) pp.61-87
[34] Brown & Rozeff (1978) pp.1-16
[35] Foster (1986) pp.279-280
[36] Collins & Hopwood (1980) pp.390-406
[37] Collins & Hopwood (1980) pp.390-406
[38] Finn & Whittred (1982) pp.171

H3) *Security Analysts make less accurate earnings forecasts for small companies than for medium and large companies*

One issues of forecasting is how to handle growth and volatile earnings. Size is often an indication of level of maturity of the company, as larger companies tend to grow slower and have more stable earnings than smaller companies. Therefore, we expect Security Analysts to make better forecasts for larger companies than for small companies. It should be noted however that this only applies if everything else is similar. Unfortunately, this is rare since there is a risk that the sample could be biased for example toward a specific industry. Previous research based on market values has shown that forecasts on large and medium sized companies are significantly more accurate than those for small companies.[39]

H4) *Security Analysts' forecasting accuracy varies between different sectors*

We expect to see differences between sectors, as the level of maturity, cyclicality and other characteristics vary between sectors. Although there is no clear evidence from prior research [40] , we expect to see differences for both between sectors as volatile earnings make forecast more difficult for both Security Analysts and Times-Series Models.

H5) *Security Analysts' earnings forecasts for the fourth quarter are less accurate than the first three*

For most companies, there are differences between the four quarters (Quarter 1, Quarter 2, Quarter 3 and Quarter 4), due to seasonality effects, and this tends to overinflate earnings in certain quarters. This should not however impact the predictability accuracy of the forecast as analysts are aware of this. This hypothesis rather is based on the accounting perspective where the fourth quarter usually is the quarter that contains the most non-recurring items. Since it is the last quarter of the year, all items not posted in previous quarters must be included.[41] As a result, the fourth quarter should be more volatile and harder to predict.

[39] Bathke, Lorek & Willinger (1989) pp.49-68
[40] Ruland (1972) pp.30-37
[41] Watts (1976) pp.81-85

H6) *Security Analysts tend to overestimate results in a bear market and underestimate results in a bull market*

In this hypothesis we will explore the assumption that Security Analysts might be cautious in their assumptions, resulting in a reverting to mean effect where their estimates are too low in a bull market and too high in a bear market. Although we have no previous support for this view, we find it to be an interesting circumstance worth exploring.

4 Methodology

This is a quantitative thesis studying quarterly earnings data on Swedish companies during the observation period ranging from first quarter 2001 until third quarter 2007. We assume that the sample is a fair depiction of the population and thus we may draw general conclusions based on the conclusions drawn from our sample. Earnings data was collected from SME Direct. The data sample was limited according to four criteria:

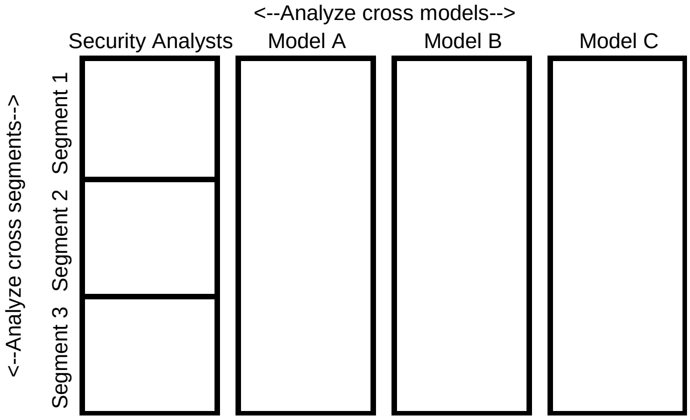
1. Company selection: All companies must be listed on the Stockholm Stock Exchange. Companies with a strong Swedish connection that are registered in other countries but have secondary listings on the Stockholm Stock Exchange, such as AstraZeneca and Nokia[42] have also been included.
2. Time Period: Only the period from Q_1 2001 to Q_3 2007 was used. A firm is, however, not required to have observations for all quarters. The relatively short time period depends on that quarterly reporting, compared to for example the US, is a relatively new phenomenon and longer time series is not available for most companies. Therefore, using a longer period would limit the availability of companies to the extent that it would render this study impossible.
3. Security Analyst Coverage Requirements: All companies must be followed by at least three Security Analysts in every measured period. If there are too few Security Analysts, there is a risk that extreme estimates will negatively influence the findings. This would mean that the Security Analysts' consensus is not representative for the opinion of the Security Analysts and it would hinder the comparability between the companies or quarters.
4. Comparability: For each company, only quarters where data is available for earnings outcome, Security Analysts forecasts and where forecasts could be generated for all of the three Times-Series Models, have been used. This is to make sure the exact same data is used in the comparison between the models. Since the Times-Series Models require different amount (and type) of historical data for its predictions, the comparison would otherwise be skewed.

From all companies on the Stockholm Stock Exchange, 74 companies (For a complete list of firms see appendix B) met our sample criteria 1-3 concerning company selection, time period and Security Analyst coverage requirements. These companies included data for 1420

[42] Note that Nokia is delisted as of June 1st 2007, but is still covered by the Security Analysts

quarters with information about earnings outcome, and Security Analysts earnings forecasts. Criteria 4, Comparability, means that every individual quarters needs to include all of the above mentioned variables and in addition calculated forecasts for each of the three Times-Series Models. After applying the fourth criteria, only 836 quarters remained. In hypotheses 1 and 2, the exact same observations (836) are used for both Security Analysts' and Times-Series Models' forecasts. This is however not the case when we look at the different circumstances (hypotheses 3-6), as the idea is to divide the 836 observations into segments. See illustration 4.1 below.

Illustration 4.1 Full sample vs. Segments



In hypotheses 6, we have expanded our sample to 1070 observations, by excluding the limitations made by Model C (which requires more historical data than the other Times-Series Models), in order to achieve a large enough sample for both circumstances (Bull and Bear).

Table 4.1 Number of observations

Hypothesis	Sample	Segments	Observ.
Hypothesis 1&2	Full Sample		836
Hypothesis 3	Size	Large	434
		Medium	177
		Small	225
Hypothesis 4	Sector	Consumer	160
		Financials	120
		Healthcare	62
		Industrials	268
		IT	100
		Materials	108
		Telecom.	18
Hypothesis 5	Quarter	Quarter 1	214
		Quarter 2	223
		Quarter 3	217
		Quarter 4	182
Hypothesis 6	Market Trend	Bull	899
		Bear	171

4.1 Reliability of Data

Earnings outcome and Security Analysts' consensus forecasts are collected from SME Direkt. This assures that earnings outcome and Security Analysts' earnings forecasts are comparable. Security Analysts consensus figures from SME are widely used by Security Analysts, investors, and media and are thus considered to be reliable. As SME doesn't make their own forecast, there is a risk that SME made a mistake when collecting data. There are, however, several institutes delivering Security Analysts numbers and the differences between them are often negligible. There is of course also a risk that we have made a mistake when transferring data into excel and SPSS. There may also be differences in the quality of Security Analysts' consensus depending on the number of participating Security Analysts. Fewer Security Analysts could result in less reliable Security Analysts consensus figures, which primarily would hit smaller companies with little coverage. We do not however know whether this would have any significant effect, but assumes that our criteria of at least three Security Analysts are sufficient.

All companies do not have data for all quarters. Most often they miss the early year of the time period but sometimes there are also quarters missing in the middle of the time series. The overall effect is that recent quarters are overweighed. We do not however feel that there that there should be a substantial difference between the beginning and end of the observation period.

Only companies that are still listed on the Stockholm Stock Exchange are included. The result is that companies that have been acquired, have defaulted or for other reasons have been delisted during the period have been excluded. Whether this has affected the data, and to what degree, is hard to determine, but as there are a large number of companies listed and only a few number of companies have left or been added to the Stockholm Stock Exchange during the period, we estimate the effect to be small. We conclude that a surviving bias is not in effect.

To conclude this section, we would like to point out that there are some difficulties in the measurement but we still believe the data is of very high quality and of a more than satisfying quantity.

4.2 Selection of Time-Series Models

We have used three Times-Series Models. Model A which is a pure Martingale model and the other two are versions of the Sub Martingale model. Two important issues when predicting earnings are how to deal with volatile earnings and growth. The three models represent different angles of how to approach these issues. Note that there are countless possible versions of Times-Series Models, and that we have only picked some of most common ones.

4.2.1 Model A

Model A is a pure Martingale model, which assumes the next result is the same as the previous comparable result.[43] [44]

$E(Q_t) = E(Q_{t-4})$

Where E(Q) is quarterly earnings and
t is the forecasted period

In the long run, earnings has to grow with at least the same pace at GDP to not lose market shares and by the inflation rate, in order to achieve a zero real growth. When results are presented in news and Security Analysts' reports, there is almost always a comparison between the current result and the previous comparable figure, as the market often expects a result of at least the same level as the previous comparable result. This model is a simple approximation and contains several sources of potential errors, such since not assuming zero growth, not dealing with volatile results and not taking non-recurring items into consideration. It should, as no growth is assumed, generally be underestimating results.

4.2.2 Model B

Model B is a Sub Martingale model based on the assumption that" earnings will revert toward previous levels after drifting temporarily".[45] The next period's earnings are assumed to be the average of the four last quarters,[46] capturing both the value of the adjacent and the comparable quarter as discussed earlier.[47]

[43] Ruland (1972) pp.30-37
[44] Foster (1977) pp.1-21
[45] Ruland (1972) pp.30-37
[46] Foster (1986) p.263
[47] Foster (1977) pp.1-21

$$E(Q_t) = \sum_{-4}^{-1} E(Q_t)/4$$

Where E(Q) is quarterly earnings and
t is the forecasted period

Compared to Model A; Model B uses an average of several quarters making it better at incorporating the problems of volatile results and effects of non-recurring items. The model does however not fully address the issue of growth as no additional growth from the average of previous quarters has been added, and therefore Model B should be underestimating results. An additional problem with pure Martingale models is that seasonal effects might distort the result. It will on the other hand improve forecasting, compared to Model A, for companies where earnings volatility is high (but not due to a seasonal pattern) from quarter to quarter.

4.2.3 Model C

Model C is a version of the Sub Martingale model with a drift factor that assumes earnings growth in the predicted period is the same as for the previous comparable period[48]:

$$E(Q_t) = E(Qt_{-4}) * (1 + (E(Q_{t-4}) / E(Q_{t-8})))$$

Where E(Q) is quarterly earnings and
t is the forecasted period

When predicting growth, the previous period's growth rate is often the starting point for the next period. Using the previous comparable quarter's growth rate may however lead to large deviations if earnings are volatile.

4.2.4 Alternative Methods

One alternative would be to use a Box-Jenkins method, which is a common tool for evaluating long time series to make prediction. The Box-Jenkins is a more advanced model that locates the autoregressive integrated moving-average (ARIMA) model which best represents the series" [49]. The Box-Jenkins method does however require long time series of data, which becomes an issue as quarterly data is a relatively new phenomenon in Sweden. Previous studies, including research by Watts & Leftwich and Albrech, Lookabill &

[48] Foster (1977) pp.1-21
[49] Ruland (1972) pp.30

McKeown on the other hand do not support that the Box-Jenkins method provides better forecasts on annual earnings than random walks models. [50] [51]

4.3 Forecast Evaluation Methods

We will use two different forecast evaluation methods: Percentage Error Analysis and Regression Line Analysis. Percentage Error Analysis is the method most commonly used in previous earnings forecast literature. We have also chosen to use a Regression Line Analysis in order to complement, control and adjust for the shortcoming of the Percentage Errors Analysis.

4.3.1 Percentage Error Analysis

We will use two versions of the percentage error: absolute percentage error (APE) and regular percentage error (RPE). Both versions use the difference between outcome and forecast in the nominator and the outcome in absolute values in the denominator. The difference between APE and RPE is that the APE uses the absolute value also in the nominator. This implies that the APE measures the percentage error unrelated to its direction (positive or negative), while the RPE could be either positive or negative. The percentage errors are calculated for each single observation.[52]

$$\text{APE} = \frac{|\text{Outcome-Forecast}|}{|\text{Outcome}|}$$

$$\text{RPE} = \frac{\text{Outcome-Forecast}}{|\text{Outcome}|}$$

APE= Absolute Percentge Error
RPE= Regualr Percentage Error

We will calculate a mean and a median for the observation sample. The difference between the mean and median RPE and APE is that APE only measure positive values while RPE has both negative and positive values which may cancel each other out. APE therefore measures the size of the mean or median error (always positive) while the RPE only tells us whether the positive values are greater in number (median) or size (mean) than the negative values. For

[50] Watts & Leftwich (1977) pp.253-271
[51] Albrecht, Lookabill & McKeown (1977) pp.226-244
[52] Foster (1977) pp.1-21

the RPE, we will also look at the number of values that are negative (overestimation) versus the number that are positive (underestimation). This will be referred to as the quantity.

4.3.2 Regression Line Analysis

The method to extrapolate the linear regression line from a cluster or observations is fairly straightforward. The line is defined as the line that cuts the sample so that the squared vertical distance from the line is minimized. Despite being simple to extrapolate, the regression line still provides ample of information. The linear regression line relies on the assumption that the residuals are normally distributed. This is often tested by plotting the residuals in a normal probability plot where the residuals should form a straight line.

Regression line

$y = b_0 + b_1 * x$

$$b_1 = \frac{\sum (x_i - X)(y_i - Y)}{\sqrt{\sum (x_i - X)^2}}$$

$b_0 = Y - b_1 * X$

$x =$ Reg. Line x value
$y =$ Reg. Line y value
$x_i =$ Individual observations x value
$y_i =$ Individual observations y value
$X =$ Average x value
$Y =$ Average y value

The r^2 denotes the match between the observations and the linear regression line. In other words it describes how well the individual observations conform to the regression line. Thus a high r^2 value denotes a high compression of the observations around the regression line. The r^2 value is between 0 and 1, where 0 is no correlation and 1 means 100% correlation.
Note that it is possible to achieve the exact same regression line given two completely different samples with two different r^2 values. The slope, b_1 value, denotes by what magnitude the line grows in relation to the two variables. The Y-intercept is the point where the line crosses the y axis. A perfect match between forecast and outcome suggests a Y-intercept of zero.

The Regression Line Analysis is used to eradicate the near zero earnings problem when using percentage error as a basis for evaluation. For each of Security Analysts and the Times-Series Models, we will prepare a regression line by plotting all the observations for its earnings forecasts and earnings outcome in a graph. We put earnings outcome on the Y-axis and forecasted earnings on the X-axis. We then calculate a linear regression line for the observations. This regression line is then compared with a perfect forecast, i.e. earnings forecasts equal earnings outcomes, which is illustrated by a straight line with a slope of 45 degree.

4.4 Method of Analyzing the Data

We will continue by explaining how we will analyze our data. The method for analyzing each hypothesis will be explained below in section 4.4.1-4.4.4.

4.4.1. Security Analysts Provide more Accurate Forecasts than Time-Series Models (H1)

To compare the forecasts between Security Analysts and the Times-Series Models we start by looking at the APE. The reason for looking at APE and not RPE is that we assume that the user of forecasts, primarily an investor, is affected to the same degree by an overestimation as by an underestimation of earnings. The forecast will affect the investors' decision, of whether to buy, hold or sell the stock ahead of the report. An investor also often has the opportunity to take both long and short positions. This means that a positive earnings surprise is only good for someone owning the share, and not for someone who has declined to buy the share, sold its shares or gone short in the stock ahead of the report. This highlights the importance of accurate forecasts and the fact that it doesn't matter if the forecasts are over- or underestimating, only the size of the error in absolute terms. One overestimation and one underestimation do not negate each other. Instead, they merely result in two errors that can lead to poor investment decisions which is why RPE should not be used.

We continue by looking at the mean APE for Security Analysts and each of the Times-Series Models. In order to draw any conclusions from this we must test the statistical significance. The most commonly used test in previous research has been the paired t-test. However, as the paired t-test relies on the assumption that the differences between the pairs are normally distributed, we first test the distribution by a Kolmogorov-Smirnov test[53]. If the distribution is normally distributed, we use a t-test to test the differences between AS and each of the Times-Series Models.

[53] See Appendix 2.1

Next, we look at the median APE, which in comparison to the mean does not take the individual size of the deviations into consideration. We use Wilcoxon Signed Rank test[54] to test the differences between Security Analysts and each of the Times-Series Models. Wilcoxon Signed Rank test does not require a normal distribution. We will furthermore compare also compare the size of Security Analysts' and Times-Series Models' first and third quartiles (the median being the second).

One of the drawbacks of using forecast error is that it is very sensitive to deviations from the outcome when the earnings outcome is close to zero. For example, if a forecast misses an outcome of SEK 1,000m by SEK 10m, it gives a deviation of 1%. If the outcome instead is SEK 0.2, a deviation of SEK 10 indicates a deviation of 5,000%.

One way of dealing with the deviations of values close to zero would be to simply exclude all outliers. According to the statistical definition, the cut off point for outliers and extreme outliers will be derived by multiplying the spread between quartile three and quartile one by 1.5 for outliers and by 3 for extreme values[55]. One of the drawback of using the statistical definition is that the cut-off point will differ between the models. Regardless of the model used, this discrepancy has an equally negative impact for investors and therefore, the cut-off point should be kept consistent across all models.
In order to make sure that the forecast error is not caused by the measurement problem with small earnings, we apply a Regression Line Analysis. The regression is less sensitive towards outliers as it doesn't measure deviation in percent but rather in absolute values. We will look at the r^2 values to determine how well the observations are gathered around the regression line, where a higher r^2 indicates a better correlation with the regression line. [56]

Using the r^2 value does however assume that the regression line equals the 45 degree regression line. We therefore need to test whether the regression line corresponds to the 45 degree line. In order to determine if the regression line and the 45 degree line are statistically the same line we perform an F test[57]. Depending on the results from the F-test, we might extend our analyses with a discussion of the practical significance. The F test and Practical

[54] See Appendix 2.2
[55] See Appendix 2.3
[56] See Appendix 2.4
[57] See Appendix 2.5

significance[58] is only used to test which samples may be compared with each other using the r^2 analysis.

In order to determine the practical significance of any differences between the individual regression line, and the 45 degree line we must first try to visually detect any significant differences. This will tell us how the difference between the lines will affect an investor. If the deviation is small we will accept that the two lines are practically the same.

4.4.2. The Time-Series Models Tend to Underestimate Earnings while Security Analysts neither Overestimate nor Underestimate Earnings (H2)

In order to determine whether Security Analysts or Times-Series Models over- or underestimate, we start by looking at the quantity (as defined in 4.3.1) for RPE. We use a Binomial-test[59] to determine if the quantity is statistically significant from a non-skewed relationship of 50% for each of over-and underestimations.

The median RPE is similar to RPE quantity as it looks at the observation in the middle of the sample. In comparison to the quantity, it does however also add the magnitude of any error. This implies that we will be able to rank the size of the over- or underestimation for Security Analysts and Times-Series Models. We use a Wilcoxon Signed Rank test to determine if results are significant.

Next, we will evaluate if the size of earnings affects if the Security Analysts or the Times-Series Models over- or underestimates. An example would be to underestimate large positive numbers while underestimating small positive numbers and negative values. For this we will use the Regression Line Analysis to compare the slopes of the regression lines. This is followed by a discussion of whether the regression lines are good proxies for the actual observations by discussing the distribution of observations and the r^2value of the slope.

[58] See Appendix 2.6
[59] See Appendix 2,7

4.4.3 Circumstances: Size, Sectors and Quarters (H3-5)

To find out whether the Security Analysts' forecast accuracy varies with different circumstances we first look each segment's median APE. Any detected differences are then tested by a Mann-Whitney U test[60] at 5% and 1% significance level.

In order to evaluate if the median APE has been influenced by the outliers or extreme outliers, we calculate the r2 value for each of the segments and evaluate any differences. In other words, we statistically evaluate the difference between the two lines utilizing an F test. Depending on the results from the F-test, we might extend our analyses with a discussion of the practical significance.

4.4.4 Circumstances: Market Trend (H6)

Similar to the approach used in H2, we use the RPE quantity to explore whether the segments over- or underestimate earnings, and test the significance with a non parametric Binomial-test.

We will continue by looking at the median RPE, in order to rank Security Analysts and the Times-Series Models according to the size of their over- and underestimations. Statistical significance will be calculated with a Wilcoxon Signed Rank test.

[60] See Appendix 2.8

5 Definitions

5.1 Definition of Earnings

There are several alternatives when measuring earnings, ranging from revenues stated at the top of the income statement to net income found at the bottom and several other measures in between. Sales is less volatile but would on the other hand miss changes in the cost structure, which would imply different growth in top- and bottom line results. Bottom line earnings, in contrast, include all items and furthermore reflect the shareholders' profit. We have, however, chosen to add back tax to look at Earnings Before Taxes (EBT) since the cost of tax is not a good indicator of the company's performance especially in a long-term perspective. There are several reasons for this, including the fact that tax is not a cost that a company can control. Similarly, corporate taxes in Sweden can be postponed (tax reserves) and brought back in later years, which does not comply with the accounting principle of matching. Companies with operations in other countries may be subject to different and varying tax rates and as a result they may to some extent also transfer costs and income between countries to affect (often to minimize) the tax rate, which do not accurately reflect company performance.

Another possible definition of earnings would be Earnings Per Share (EPS), which is the definition of earnings that has been most frequently used in previous similar studies. There are, however, several drawbacks when using a ratio like EPS. Firstly, the difference in EPS between two years is not only dependant on the difference in earnings but also on the changes in the number of shares. The number of shares could change as a result of new share issues, splits or share buy backs, which in turn would affect EPS, despite not being a measurement of the company's performance. Secondly, EPS is often fairly small (sometimes only a couple of Swedish öre) and is only measured by two decimals of a Swedish Krona (as öre is the smallest unit of measurement), which leads to a measurement problem. A change from for example SEK 0.01 to SEK 0.02 would indicate a growth of 100%. If using more decimals, the difference could range from slightly more than 0% (SEK 1.499… vs. SEK 1.500…) to slightly less than 400% (SEK 0.501… vs. SEK 2.499…). A third problem arises when creating a regression. EPS gives the same percentage deviations. However, a regression does not look at percentage deviations but deviations in absolute numbers. This means that the impact of observations with the same deviations could have different impact on the regression line depending on the size of the absolute number of the EPS, which in turn is a result of the number of shares. EPS appears to offer no real advantage other than the fact that it is commonly used, therefore, we have opted to use EBT as our measurement and will hereafter use the term “earnings” in reference to EBT.

5.2 Currency

All earnings numbers are measured in Swedish Krona. Earnings from companies reporting in other currencies have been translated into SEK. Since forecasts in these cases are made in the same currency, there should be no discrepancy between outcome and forecast. The implication if not using the same currency, is that it will be misplaced on the regression line. We have measured the currency annually on January 1st. The exchange rate has thereafter been used for all four quarters of that year. All earnings figures presented are measured in millions of Swedish Krona (MSEK).[61] (The Exchange rates used are found in Appendix C)

5.3 Quarterly earnings

Previous studies on quarterly earnings have found evidence that there is both a seasonal component (with the previous comparable quarter), and an adjacent component (with the previous quarter). The seasonal component has larger descriptive validity, but incorporating both components a give higher descriptive validity than only using one of them. [62]

With support by from the aforementioned studies, we have used comparable quarters when creating forecasts with our models. This goes for all models except model B (see 5.1.2 for model B), where we have used the four previous quarters. In addition, using only comparable quarters would demand too many previous comparable quarters and thereby make the time series too short.

5.4 Definition of Consensus Forecasts

We have defined consensus as a group of at least three Security Analysts. The largest number of analyst in our sample was 48 and we had an average of 12.8 Security Analysts per consensus forecast. The Security Analysts figures used are the ones reported closest to the release of the report and this data is usually collected. This data is usually collected by specific news related companies over the phone a couple of days before the reporting date.

---

[61] For Exchange rates see Appendix 3,2
[62] Foster (1977) pp.1-21

We have furthermore, opted to use the median consensus forecast in order to exclude the effect of any extreme forecasts. When looking at our results it is apparent that the consensus median and the consensus mean are extremely similar and therefore is of minor importance for our results.

5.5 Circumstances

5.5.1 Definition of Size

The influence of different firm sizes on the ability to accurately forecast earnings was addressed by Brown, Richardson & Schweger in 1987. However, it should be noted that there are several ways of measuring size including sales, earnings, assets, number of employees and market cap. [63]We have opted to follow the same methodology as Bathke, Lorek, Wilinger in their study on US companies. In accordance with this study we have defined size as market capitalization and have consequently divided the companies into three groups: Small, Mid and Large.[64] As the relative size of Swedish and US companies differ, we have used our own definition of the three size groups, and defined Small as companies with a market cap below SEK 10bn, Mid as companies with a market cap of SEK 10-30bn and Large as companies with a market cap above 30bn. This type of classification or division would be considered a common industry practice.

Market capitalization, as all other measures, varies over time. Since the same companies should be included in the same size group for the entire time-series, we have based the market cap on the most recent date available (6th of February 2008). One adjustment to the size groups has been made, as we have moved OMX from Large to Medium, despite its market cap of SEK 31.8bn. This was done because historically, OMX has had a lower valuation and was only valued higher in 2007 as a result of a bid on the company.

5.5.2 Definition of Sectors

We have defined our sectors according to the Global Industry Classification Standard (GICS). This is a generally accepted standard which is widely used in Sweden as well as abroad. It is furthermore used by the OMX Nordic Exchange . GICS recognizes 10 sectors, with 9 of the sectors represented on the Stockholm Stock exchange; Energy, Materials, Industrials,

[63] Brown, Richardson & Schweger (1987) pp.49-67
[64] Bathke, Lorek & Willinger (1984) pp.49-68

Consumer Discretionary, Consumer Staples, Health Care, Financials, Information Technology and Telecommunication Services (Telecom). [65]We have made some further adjustments to this and we have moved Lundin Petroleum from Energy to Materials, since Lundin Petroleum was the only company in the Energy sector. Lundin Petroleum is also, like the companies in the materials sector, a resource company. We have furthermore merged Consumer Discretionary and Consumer Staples, to get a larger sample and feel that this is appropriate since the definition for both sectors are quite vague. We have also moved Intrum Justitia from Industrials to Financials. The logic behind this is that its business closer resembles the businesses in the financial sector and is usually also covered by the same Security Analysts as those covering the financial sector.

5.5.3 Definition of Quarters

The fourth quarter for a company is defined as its fiscal fourth quarter. The reason for this is that we are primarily looking for an accounting effect, and not a seasonal effect. Most companies however have fiscal year ends that are consistent with the calendar year so the difference is minimal.

5.5.4 Definition of Market Trend

We have divided our time series in two parts. Each series represents a market trend based on the stock market's performance. The first period, from Q_1 2001 to Q_1 2003, is a period of decline in the stock market and will be referred to as Bear. The stock market changed direction during the spring of 2003, after reaching its lowest level on the 31st of March 2003 and has trended up to the end of our observation period. We will refer to the period of Q_2 2003 to Q_3 2007 as Bull. This results in a time series that is twice as long for Bull.

[65] OMX Nordic Exchange <http://omxnordicexchange.com/forbolagochemittenter/produkterochtjanster/Sektorindelning>/4th May 2008

6. Empirical results

6.1 Observations and Distribution

As described earlier, the full sample consists of 836 observations. The sample is then divided into several segments dependant on size, sector, quarters and market trend. The full sample is expanded to 1,070 observations when analyzing the market trend (H6). This is done by removing the restriction caused by Model C. We see that there are several differences, not only between the Security Analysts and the Times-Series Models at the full sample, but also when segmenting the sample.

When looking at the size circumstance it is worth noticing that a larger size of the firms has not only a higher median but also a wider range. To illustrate, Large has a spread from -3,407 to 23,102 compared to Small that only has spread from -1,583 to 969.

When looking at the differences between the quarters we find that the only quarter that exhibits a different characteristic is the Quarter 2 that only has a minimum value of -532 while the other quarters bottom at roughly -3,500. All quarters have the same maximum earnings level of approximately 2,000.

When looking at the sectors, it gets more complicated as there is a larger spread between the different sectors. We can see that the lowest as well as the highest earnings comes from an IT firm. It is also interesting to observe the performance of Healthcare firms. Healthcare is the only segment that has no losses (negative earnings) while being in the top when it comes to maximum profits.

Bull and Bear have similar numbers, although Bull has higher maximum, third quartile and mean.

Table 6.1 Distribution of earnings outcome (MSEK)

Sample	Minimum	1st Quartile	2nd Quartile	3rd Quartile	Maximum	Mean	Median
Full sample	-3 407	173	579	1 973	23 102	1 756	579
Large	-3 407	870	1 828	3 870	23 102	3 132	1 828
Mid	-659	264	403	622	2 364	487	403
Small	-1 583	25	84	157	969	100	84
Quarter 1	-3 407	158	581	1 870	14 673	1 692	581
Quarter 2	-532	202	579	2 013	23 102	1 858	579
Quarter 3	-3 396	169	610	1 853	18 171	1 673	610
Quarter 4	-3 229	162	558	2 034	17 402	1 802	558
Consumer	-237	161	464	898	5 440	809	464
Financials	-532	214	2 526	3 866	9 605	2 571	2 526
Healthcare	29	253	411	7 688	15 120	3 680	411
Industrials	-1 583	276	801	1 660	6 673	1 206	801
IT	-3 407	17	231	8 439	23 102	3 702	231
Materials	-3 396	153	572	1 604	2 898	801	572
Telecom	-2 697	235	655	1 666	5 932	1 194	655
Bull	-3 407	153	581	1 456	11 447	1 276	581
Bear	-3 396	173	578	2 006	23 102	1 773	578

6.2 Mean Absolute Percentage Error

The mean APE for the full sample ranges from 34.0% for Security Analysts to 601.7% for Model C. Model B (84.3%) has the lowest mean APE of the Times-Series Models. When dividing Security Analysts into different segments we get a range of 12.6% for Healthcare to 78.1% for Small. The largest difference between the highest and lowest mean APE within a circumstance is in Size (61.2 percentage points) and the smallest in Market Trend (0.6% percentage points). In table 6.3 you can also see the mean APE without outliers, mean APE without extreme values, and the maximum value for APE (the single highest observed APE).

Table 6.2 Mean Absolute Percentage Error

Sample	Segments	SA	Model A	Model B	Model C
Full Sample	Full sample	34,0%	95,0%	84,3%	601,7%
Size	Large	16,9%			
	Medium	20,1%			
	Small	78,1%			
Sector	Consumer	22,7%			
	Financials	18,3%			
	Healthcare	12,6%			
	Industrials	39,8%			
	IT	68,2%			
	Materials	31,9%			
	Telecom.	50,3%			
Quarter	One	24,0%			
	Two	41,7%			
	Three	33,4%			
	Four	37,1%			
Markettrend	Bull	32,5%			
	Bear	33,1%			

Table 6.3 Mean Absolute Percentage Error

	Security Analysts	Model A	Model B	Model C
Mean	34,0%	95,0%	84,3%	601,7%
Mean Without Outliers	10,2%	32,4%	29,6%	50,2%
Mean without Extreame values	12,0%	38,7%	35,3%	58,5%
Maximum Values	2050,0%	5850,0%	5050,0%	305250,0%

6.3. Median Absolute Percentage Error

The median APE for the full sample range from 9.2% for Security Analysts to 40.1% for Model C. Model B (23.8%) has the lowest median APE of the Times-Series Models. When dividing Security Analysts into different segments we get a range of 3.9% for Healthcare to 16.7% for Small. The largest difference between the highest and lowest median APE within a circumstance is in Sector (11.2 percentage points) and the smallest in Quarter (0.9% percentage points).

Table 6.4 Median Absolute Percentage Error

Sample	Segments	SA	Model A	Model B	Model C
Full Sample	Full sample	9,2%	27,8%	23,8%	40,1%
Size	Large	7,3%			
	Medium	7,1%			
	Small	16,7%			
Sector	Consumer	7,7%			
	Financials	9,0%			
	Healthcare	3,9%			
	Industrials	9,8%			
	IT	12,5%			
	Materials	13,7%			
	Telecom.	15,1%			
Quarter	One	8,9%			
	Two	9,6%			
	Three	9,0%			
	Four	9,8%			
Markettrend	Bull	9,3%			
	Bear	10,3%			

6.4 Median Regular Percentage Error

The median RPE for the full sample range from -2.4% for Model C to 12.8% for Model A. Security Analysts have a median RPE of 2.8%. When looking at Market Trend (the only circumstance where we use median RPE), Bear has a value of 3.23% and Bull 2.59.

Table 6.5 Median Regular Percentage Error

Sample	Segments	SA	Model A	Model B	Model C
Full Sample	Full sample	2,8%	12,8%	9,6%	-2,4%
Markettrend	Bull	2.6%			
	Bear	3,2%			

Due to the way RPE is defined, a positive RPE indicate an underestimation and vice versa. Model C has the lowest deviation from a random distribution with 49.4%, indicating a slight tendency to overestimation. While Model A underestimates the most off all Time-Series Models and Security Analysts (67.1% of the observations).

Table 6.6 Over- & Underestimations

Model	Segment	Observ.	Underest.
Security Analysts	Full Sample	827	59.3%
Model A	Full Sample	833	67.1%
Model B	Full Sample	836	62.3%
Model C	Full Sample	836	49.4%
Security Analysts	Bear	170	58.8%
Security Analysts	Bull	884	58.2%

6.5 Regression Line

Security Analysts (-2.8) have the lowest Y-intercept (measured in millions of SEK) and Model C the highest (1681.5). Security Analysts also have the value closest to zero. Model B (115.1) is the Times-Series Model with the value closest to zero. Model C (0.390) has the lowest slope and Security Analysts have the highest (1.054). Model A (1.000) has the slope closest to 1. Security Analysts (0.976) have the highest r^2 and Model C the lowest (0.027). Model B (0.907) is the Times-Series Model with the highest r^2.

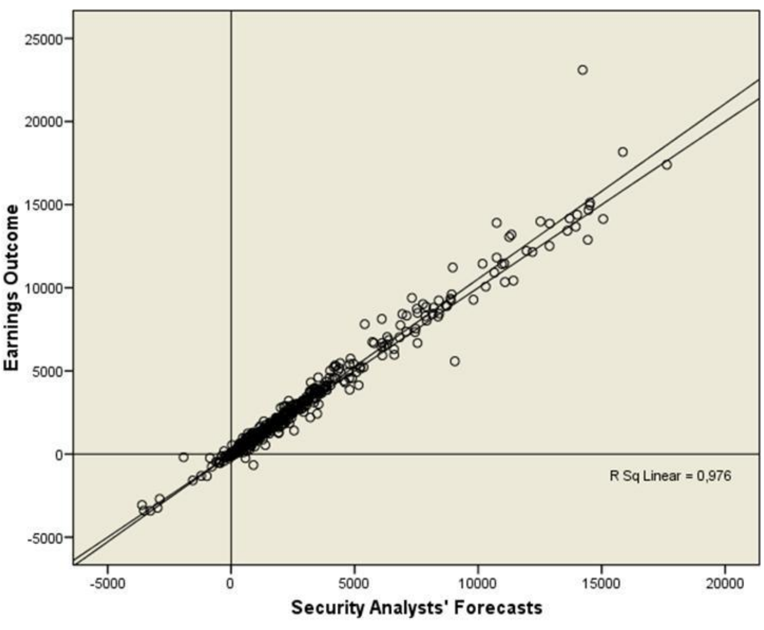
The Y-intercepts for the segments range from -26.6 for Quarter 1 to 106.2 for IT. Materials (2.8) has the value closest to zero. All segments subsequently have intercept further away from the full sample. The slopes range from 0.913 for Mid to 1.088 for Quarter 2. Materials (0.991) has the slope closest to 1. Quarter 4 (0.989) has the highest r^2 and Mid and Industrials (0.858) the lowest.

Table 6.7 Regression Lines

	Y-intercept	Slope	R2 Value
Full Sample:			
Consensus	-2,8	1,054	0,976
Model A	284,0	1,000	0,841
Model B	115,1	1,033	0,907
Model C	1 681,5	0,390	0,027
Segments:			
Large	30,8	1,050	0,970
Medium	34,2	0,913	0,858
Small	-10,3	0,972	0,907
One	-3,2	1,074	0,988
Two	-26,6	1,088	0,952
Three	10,5	1,038	0,983
Four	-5,1	1,019	0,989
Consumer	-18,0	0,980	0,987
Financials	63,8	0,897	0,985
Healthcare	4,5	0,947	0,988
Industrials	45,5	0,931	0,858
IT	106,2	0,911	0,962
Materials	-2,8	0,991	0,957
Telecom.	35,9	1,061	0,985
Bull	33,1	0,926	0,988
Bear	39,9	0,929	0,976

In Graph 6.1 we can see one visual example of how a regression line and its plotted observations look like. This example is for Security Analysts (full sample). The information this graph tell us about the distribution of the observations, the gathering of observations around the regression line and how close the regression line is to the 45 degree line.

Graph 6.1 Scattergram of Security Analysts' earnigs forecasts and Earnings outcome



7 Analysis

7.1 Security Analysts provide more accurate forecasts than Time Series Models (H1)

7.1.1 Percentage Error Analyses

When looking at the mean APE, we can see that Security Analysts has the lowest mean of 34.0%, while the best Times-Series Models (model B) has a mean of 84.3%, which is almost 2.5 times higher than that for Security Analysts. Since the APE values seem fairly high, we look at the maximum deviation, where Security Analysts have a significantly lower value (2,075%) than the Times-Series Models (Model B is the best Times-Series Model with a maximum deviation of 5,050%). If excluding outliers and extreme outliers, the relation between Security Analysts and the Times-Series Models seem to increase further, to factor of approximately 2.9 times. This is furthermore according to the statistical definition of outliers and extreme outliers, implying that different cut off values are used. Having the same cut-off values for Security Analysts and the Times-Series Models would improve Security Analysts performance versus the Times-Series Models.

Table 7.1 Mean Absolute Percentage Error

	Security Analysts	Model A	Model B	Model C
Mean	34,0%	95,0%	84,3%	601,7%
Mean Without Outliers	10,2%	32,4%	29,6%	50,2%
Mean without Extreame values	12,0%	38,7%	35,3%	58,5%
Maximum Values	2050,0%	5850,0%	5050,0%	305250,0%

Since the K-S test shows that that the APE observations are not normally distributed, we are unable to use a t-test. We also tested the paired difference for APE between Security Analysts and each of the Times-Series Models and found that none of the differences were normally distributed at 1% significance level[66] and we can therefore not use a paired t-test either. Although the seemingly superior mean absolute percentage error provided by Security Analysts is an indication of better performance, it could not be statistically tested either by a t-test or the paired t-test, due to their lack of normality. As a result, we will exclude the analyses of the mean APE in hypotheses 3-5 (if interested, the values could be found in Table 6.2).

[66] See Appendix Table A4.1

The median APE is significantly lower than the median APE for Security Analysts and the Times-Series Models, implying that extreme values have a large impact on the mean APE. Security Analysts still provide the best accuracy with a median APE of 9.2%, and Model B is still the best Times-Series Model with a median APE of 23.8%, which is 2.6 times higher than for Security Analysts.

Table 7.2 Median Absolute Percentage Error

	Median APE			Z	
Security Analysts - Model A	9,2%	<	27,8%	-18,9	**
Security Analysts - Model B	9,2%	<	23,8%	-18,1	**
Security Analysts - Model C	9,2%	<	40,1%	-20,3	**
Model A-Model B	27,8%	>	23,8%	-4,78	**
Model A-Model C	27,8%	<	40,1%	-8,41	**
Model B-Model C	23,8%	<	40,1%	-10,5	**

*=sign. at 5%
**=sign. at 1%

All differences in median APE between Security Analysts and all of the Times-Series Models are significantly different from each other at 1% significance level, and we may thus rank them accordingly. This shows that Security Analysts make the most accurate forecasts, followed by Model B, model A, and at last model C.

7.1.2 Regression Line Analyses

We have prepared a histogram that indicated an approximate fit to the desired bell shape, and thus assume from this point on that all residuals are normally distributed. The r^2 values clearly indicate that Security Analysts' forecasts are much tighter distributed around the regression line than for all of the Times-Series Models. Security Analysts have an r^2 value of 0.976 while the most accurate Times-Series Model (Model B) has an r^2 value of 0.907. We can also see that model C has an r^2 value of 0.027, indicating an almost non-existing correlation. The low r^2 value is a result of a few extremely deviating observations, making the regression line not representative for the rest of the observations. [67] Although not possible to test for, the differences between Security Analysts' r^2 values and the Times-Series Models' r^2 values are so large that may be assumed not to be a result of measurement errors. The relatively small deviations from the regression line indicated by the r^2 value show that low earnings don't influence the conclusions derived from the median APE.

[67] For Graph see Appendix 3.3

Table 7.3 Regression line Full Sample

Forecast	Observ.	Y-intercept	Slope	R2 Value	F-value	
Security Analysts	836	-2,8	1,054	0,976	58,97	**
Model A	836	284,0	1,000	0,841	23,77	**
Model B	836	115,1	1,033	0,907	18,48	**
Model C	836	1681,5	0,390	0,027	7130,22	**

*=sign. at 5%
**=sign. at 1%

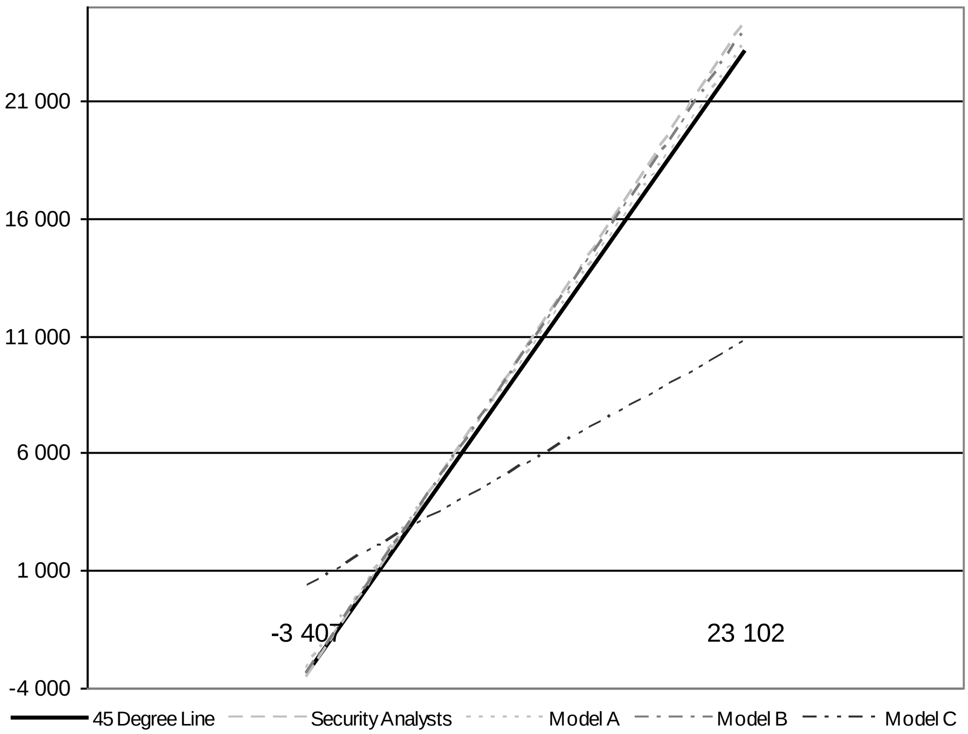
According to the F-tests, Security Analysts and all of Times-Series Models are rejected as being identical to the 45 degree line at 1% significance level. It should be noted that the Times-Series Models provided a better F-statistic than Security Analysts. This doesn't, however, indicate that they are a better approximation, but rather that they are harder to reject as their observations are more scattered (indicated by their lower r^2 value) than Security Analysts. The main reason why all models are rejected is because the large number of observations in our sample creates a situation where even small and potentially non-significant deviations cause the null hypotheses to be rejected. Although the differences are statistically significant, the differences may not be practically significant for a user of the produced forecasts.

We can see that Security Analysts' regression line and the 45 degree line are close to each other. When taking the distribution of observation into account, we can see that the clear bulk of the observations are in an area of the graph where the two lines are very close. This would support that the effect from any statistical differences between the two lines are relatively small to an investor. We therefore conclude that the two lines are practically the same as each other. You can see the two lines and the observations in a scatter gram (Graph 6.1).

Model A's regression line has a slope of 45 degrees, but has a Y intercept that negatively influences its correlation with the 45 degree line. We can see that the two lines are close to each other, which indicate that the regression line may be approximately used as the 45 degree. Model B's regression line has an even better fit with the 45 degree line than Model A, and can therefore also be assumed to be a good proxy for the 45 degree line.

When comparing the regression line for Model C with the 45 degree line we can clearly see a difference. We can therefore conclude that the regression line is not a fair proxy of the desired 45 degree line, and the r2 value can thus not be used in our analysis.

Graph 7.1 Regression Line Full Sample



7.1.3 Further Analysis

Our findings that Security Analysts provide more accurate forecasts than Times-Series Models are not only in line with our predictions but also with previous research (Brown, Griffin, Hagerman & Zmijewski in 1984 and Brown & Rozeff in 1978. [68] [69] Security Analysts have the lowest median, first and third quartiles and the lowest maximum deviation, which is in line with Elton, Gruber & Gultekin and Collins & Hopwood's findings that Security Analysts are superior since they have the ability to incorporate several different variables and incorporate new information as it becomes available. [70] [71]

From an investor's perspective, the accuracy of forecasts is extremely important. The gain in return on investments by using Security Analysts' forecast compared to Times-Series Models' forecasts ought to be extremely large. Accurate forecasts are important also in a long-term perspective when valuing companies. When looking at the widely used multiple valuation, the percentage error will incorrectly value the company with the same percentage. Also, when dealing with fundamental valuation, forecast accuracy is important.[72]As growth is most often

[68] Brown et.al. (1984) pp.61-84
[69] Brown & Rozeff (1978) pp.1-16
[70] Elton, Gruber & Gultekin(1984)pp.351-363
[71] Collins & Hopwood (1980) pp.390-406
[72] Brown & Rozeff (1978)pp.1-16

most based on a percentage number on the previous period, one incorrect earnings forecast will affect the forecast for all the subsequent periods and the size of the steady state value, and forecast will thus have large ramifications on the valuation of the company. Since all investors basically already rely on Security Analysts' forecasts (consensus or individual), it is essential to have forecasts that are superior to the consensus, which has proven to be difficult, but is necessary in order to earn abnormal earnings. On an aggregated level, our results show that an investor needs to be able to make a smaller error (measured as median) than 9.2% (median APE for Security Analysts) in order to beat consensus and thereby earn abnormal return. We conclude that the superior performance of Security Analysts to Times-Series Models is large enough to warrant the higher cost of Security Analysts compared to the relatively low cost of Times-Series Models.

Although Security Analysts' forecasts have proven to be superior to Times-Series Models, we believe there is still value in and demand for Times-Series Models. We believe that the value of Times-Series Models is rather an inexpensive way of providing a basis for the Security Analysts' forecasts. This is to a large extent already the case. [73] We do however see a risk for Security Analysts in using Times-Series Models forecasts as a basis, since Times-Series Models rely too heavily on a trend. This could result in Security Analysts becoming less attentive in detecting shifts in trends. It should also be noted that it is possible to make more complicated time-series models that incorporate more variables (multivariate Time-Series Models). It is also possible to adjust the Time-Series Models to a specific sector or even to individual companies, which would likely improve its performance. We don't, however, believe they will be able to perform as good as or better than Security Analysts.

The superiority of Sub Martingale (Model B) over the Martingale (Model A) is in line with previous research.[74] This shows that a drift factor, at least partly manages to incorporate the growth in earnings due to reinvested earnings and inflation. The poor performance by Model C does, however, show that all Sub Martingale models don't outperform the Martingale model. Model C's low forecasting accuracy highlights the implications of assuming the same growth as in the previous period, which is a commonly assumption (or at least a starting point) when forecasting growth.

[73] McEnally (1971)pp.687-706
[74] Ball & Watts (1972) pp.663-682

7.2. The Time Series Models Tend to Underestimate Earnings while Security Analysts neither Overestimate nor Underestimate Earnings (H2)

7.2.1. Percentage Error Analyses

The RPE quantities show that Security Analysts, Model A and Model B underestimate earnings at 59.3%, 67.1% and 62.3% of the observations respectively. These results are statistically different from 50% at 1% significance. Model C tends to underestimate a slight majority (50.6%) of the observations, and could not be rejected from 50% with 1% significance.

Table 7.4 Regular Percentage Error Quantity

Model	Observ.	Underestimations			Z	
Security Analysts	827	59,3%	>	50,0%	5.32	**
Model A	833	67,1%	>	50,0%	9.87	**
Model B	836	62,3%	>	50,0%	7.12	**
Model C	836	49,4%	<	50,0%	0.35	

*=sign. at 5%
**=sign. at 1%

Model C has the lowest median RPE of -0.3 followed by Security Analysts (2.8%), Model B (9.6%) and Model A (12.8%). Security Analysts' and the Times-Series Models' median RPE are significantly different from each other at 1% significance level, except for Model C and Security Analysts where the significance is only 5%. We may therefore conclude that the ranking is statistically significant.

Table 7.5 Regular Percentage Error

	RPE Median			Z	
SA-Model A	2,8%	<	12,8%	-8,1	**
SA-Model B	2,8%	<	9,6%	-4,99	**
SA-Model C	2,8%	>	-0,3%	-2,38	*
Model A-Model B	12,8%	>	9,6%	-4,2	**
Model A-Model C	12,8%	>	-0,3%	-8,3	**
Model B-Model C	9,6%	>	-0,3%	-5,2	**

*=sign. at 5%
**=sign. at 1%
SA= Security Analysts

7.2.2. Regression Line Analysis

In order to decide whether the forecasts are affected by the size of earnings, we turn to our regression analysis that measures the systematic errors. Model A has a slope of exactly 1.000

and therefore shows no systematic bias toward the size of earnings. However, as its Y-intercept is positive, it systematically tends to underestimate for all observations. Security Analysts and Model B have slopes above 1.000 and thereby, as a result, tend to underestimate more when earnings grow larger. Model C has a regression line that renders it useless for comparison. The effect from size of earnings is most easily accessible by looking at graph 7.1 found in hypothesis 1.

Next, we look at whether the regression lines are good proxies for the actual observations by looking at the r^2 values (see Table 6.5). Security Analysts show an extremely high r^2 value of 0.976 indicating an extremely high explanatory degree by the regression line. Model A and Model B also show high r^2values of 0.841 and 0.907 respectively. Although not statistically tested the difference between the models r^2 values are drastic enough to indicate that they are significantly different, we can conclude that Security Analysts experiences the highest explanatory degree by the regression followed Model B and Model A. Model C has a very low explanatory level rendering it useless.

7.2.3 Further Analysis

Our conclusions regarding the fact that the Times-Series Models underestimate results are consistent with previous research.[75]This is not surprising as both models are conservative when it comes to growth. In order to make a Times-Series Model that underestimates less, one could adjust for growth by increasing the growth in the drift factor. It comes as a surprise that Security Analysts also underestimate earnings, which should be corrected by competition between Security Analysts to produce the most accurate forecasts. Although the consensus of the Security Analysts is often seen as the market expectations, it should be noted that this is not always the case, as investors may have other earnings expectations, sometimes referred to as a “whispering number”. Since Security Analysts are aware of this, it implies that it is somehow preferable for Security Analysts to underestimate rather than overestimate. One reason for this could be that most of the Security Analysts’ recommendations are buy recommendations. Therefore it is better to have conservative forecasts in order not to disappoint investors. Another possible reason for Security Analysts’ underestimations is because of the other variable in the percentage error equation, namely the earnings outcome. Managers may have an incentive to beat Security Analysts’ forecast. Since managers have

[75] Finn & Whittred (1982) pp.169-173

access to consensus figures, they have the possibility to influence the earnings results in order to exceed the consensus forecast.

Model A has a slope equal to exactly 1.000. This has nothing to do with the construction of the model but a pure coincident. When looking at the regression lines for Model A for the different sizes[76], you see that all of them have slopes below 1.0 (Large 0.955, Mid 0.722 and Small 0.741).

One reason for the SE's tendency to underestimate for larger earnings but overestimate smaller earnings could be that it may be more common to make large absolute negative deviations than vice versa. The reason for this is that income is easier to estimate, while expenses, in form of non-recurring costs, more often are unexpected.

7.3 Security Analysts Make Less Accurate Earnings Forecasts for Small Companies than for Medium and Large Companies (H3)

7.3.1 Percentage Error Analysis

Small has larger median APE (16.7%) than both Mid (7.1%) and Large (7.3%). As the differences are statistically significant at 1% significance level, we can conclude that earnings for Small are more difficult (in terms of lower forecasting accuracy) to forecast than for Mid and Large.

Table 7.6 Median APE Size

	Median APE		Z	
Small vs. Mid	16,7% >	7,1%	-6,11	**
Small vs. Large	16,7% >	7,3%	-7,50	**
Mid vs. Large	7,1% <	7,3%	-0,22	

*=sign. at 5%

**=sign. at 1%

The difference between Mid and Large is only 0.28 percentage points and therefore not statistically significant. This implies that forecasting accuracy does not improve as size grows, but rather that it is a problem only for Small.

[76] See Appendix Table A4.3

The higher median absolute forecasting error for Small is in line with the findings by.[77] Small companies do often produce small earnings. In hypotheses 1, we reach the conclusion that Security Analysts are not more affected by extreme values due to earnings close to zero than the Times-Series Models. In hypotheses 1, we don't make any conclusion whether it is a problem within the Security Analysts sample. Therefore, we extend our analyses by making a regression analyses in order to determine whether an overrepresentation of earnings close to zero might have affected the median absolute percentage error negatively.

7.3.2 Regression Line Analysis

Mid has the lowest r^2 value (0.858) followed by Small (0.907) and Large (0.970) (see Table 6.5). This is unexpected as this indicates that Mid are the hardest to forecast. Although we have not tested the differences in r^2values we assume that the difference between Large and Small f 0.970 and 0.907 is significant just as the difference between Small's and Mid's r^2 values (0.907 versus 0.858) is statically significant. Large's regression line is statistically different from the 45 degree line at 1% significance. However, both Small and Mid can't be rejected at 1% significance (see table 7.7), and their regression lines may therefore be approximated to the 45 degree line. Subsequently, as Small has a higher r^2value than Mid, we conclude that Small's median APE seems to have been biased by a lower level of earnings. It is, however, hard to weigh the results from the median absolute percentage and the results from the regression. Our conclusion is therefore that we can't say that Small is harder to forecast, as the uncertainty of the effect of earnings close to zero is too high.

Table 7.7 Regression Size

Firm size	Observ.	Y-intercept	Slope	R2 Value	F-value	
Large	434	30,8	1,050	0,970	32,94	**
Medium	177	34,2	0,913	0,858	5,12	*
Small	225	-10,3	0,972	0,907	3,96	*

As we use different samples for the three size segments, it could not, however, be ruled out that the samples are biased, for example the split between sectors. Also, the Security Analysts are not the same persons for all companies and all time periods, but depend on such factors as sector, size and time period. The above mentioned disclaimers are applicable also for the following hypotheses (H4-6).

[77] Bathke, Lorek & Willinger (1989) pp.49-68

7.4. Security Analysts' forecasting Accuracy Varies between Different Sectors (H4)

7.4.1 Percentage Error Analysis

When comparing the median APE of the different sectors we find that Healthcare has the lowest APE (3.9%) and Materials (15.1%) the highest. When comparing all sectors against each other (21 comparisons) we find that 10 comparisons are significantly different from each other. We can thus conclude that there are differences in the forecast accuracy between different sectors. In order to be able to rank a sector, we need to show that the differences between the sectors are statistically significant. Healthcare is statistically significant against all other sectors and is thereby the easiest sector to forecast. The four remaining significant differences come from Consumer that is statistically different from Industrials, IT and Materials, and from Financials that's statistically different more accurate) than IT. Note that as the significance does not only look at median, but also the distribution, it doesn't have to be easier to get significance against the highest or lowest values. This is the case with for example Telecom (15.1%), that has the highest median APE, but still only is statistically separated from one sector (Healthcare).

Healthcare (3.9%), Consumer (7.8%) and Financials (9.9%) have the lowest median APE. This comes as no surprise since both Healthcare and Consumer are non-cyclical sectors with stable earnings. Financials is dominated by the large banks (mature industry) that have modest growth and have high analyst coverage. Telecom (15.1%), Materials (13.7%) and IT (13.7%). IT is a new and developing sector with volatile earnings and with several very small companies that may yield a higher APE. Materials is a play on volumes and fluctuating commodity prices, making it difficult to forecast. The high APE for Telecom comes more as surprise as it consists of large companies in a rather non-cyclical sector. It could, however, be biased by the low number (18) of observations (still statistically significant) or due to several values close to zero. Industrials has a value in the middle of the seven segments and is hard to draw any conclusions from.

Table 7.8 Median SPE Sectors

			Median APE			Z	
Consumer	vs	Financials	7,7%	<	9,0%	-1,20	
Consumer	vs	Healthcare	7,7%	>	3,9%	-2,49	*
Consumer	vs	Industrials	7,7%	<	9,8%	-2,06	*
Consumer	vs	IT	7,7%	<	12,5%	-2,95	**
Consumer	vs	Materials	7,7%	<	13,7%	-2,53	*
Consumer	vs	Telecom.	7,7%	<	15,1%	-1,76	
Financials	vs	Healthcare	9,0%	>	3,9%	-3,36	**
Financials	vs	Industrials	9,0%	<	9,8%	-0,66	
Financials	vs	IT	9,0%	<	12,5%	-2,67	*
Financials	vs	Materials	9,0%	<	13,7%	-1,89	
Financials	vs	Telecom.	9,0%	<	15,1%	-1,49	
Healthcare	vs	Industrials	3,9%	<	9,8%	-4,18	**
Healthcare	vs	IT	3,9%	<	12,5%	-4,38	**
Healthcare	vs	Materials	3,9%	<	13,7%	-19,5	**
Healthcare	vs	Telecom.	3,9%	<	15,1%	-2,78	**
Industrials	vs	IT	9,8%	<	12,5%	-1,69	
Industrials	vs	Materials	9,8%	<	13,7%	-1,38	
Industrials	vs	Telecom.	9,8%	<	15,1%	-1,2	
IT	vs	Materials	12,5%	<	13,7%	-0,35	
IT	vs	Telecom.	12,5%	<	15,1%	-0,34	
Materials	vs	Telecom.	13,7%	<	15,1%	-0,58	

*=sign. at 5%

**=sign. at 1%

7.4.2. Regression Line Analysis

When comparing the r^2 values of the different sectors, we find that all but Industrials have undistinguishable high values ranging from 0.957 to 0.987. This implies that the large APE values have to be made at small earnings observations. Industrials is the only sector with a significantly lower r^2 value (0.858). This is probably because Industrials is a sector with large companies with large earnings. Since the r^2 value measures deviations in absolute numbers, a relatively small deviation for large earnings will still have a large impact on the r^2 value. This is not too surprising as Industrials contains several cyclical companies.

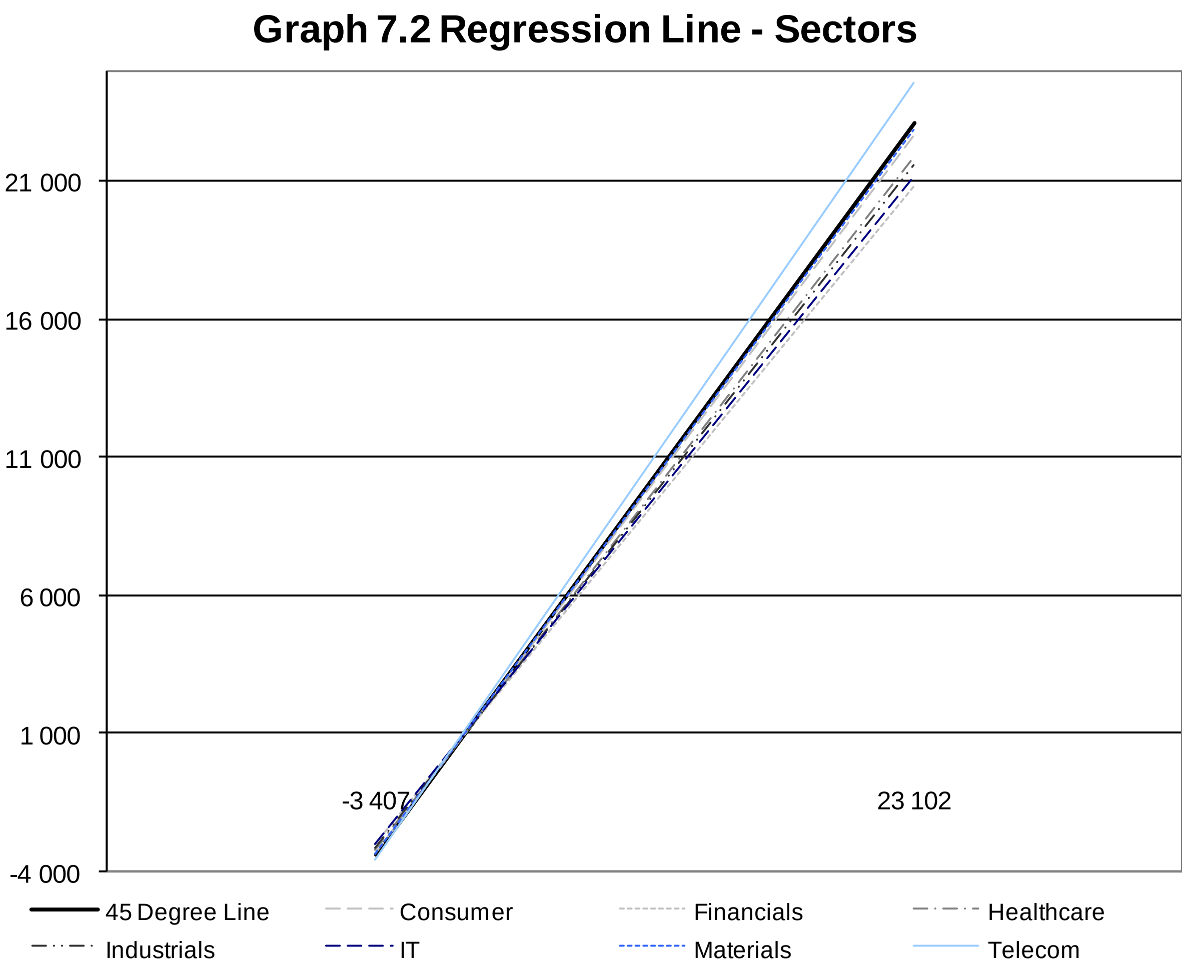
Table 7.8 Regression Size

Industry	Observ.	Y-intercept	Slope	R2 Value	F-value	
Consumer	160	-18,0	0,980	0,987	8,68	**
Financials	120	63,8	0,897	0,985	80,05	**
Healthcare	62	4,5	0,947	0,988	11,23	**
Industrials	268	45,5	0,931	0,858	26,60	**
IT	100	106,2	0,911	0,962	13,94	**
Materials	108	-2,8	0,991	0,957	0,21	
Telecom.	18	35,9	1,061	0,985	3,37	

*=sign. at 5%

**=sign. at 1%

Telecom and Materials are statistically identical to the 45 degree line at both 5 and 1% significance level. All other sectors have regression lines that are statistically different from the 45 degree line. This brings us to the question whether the statistical difference is of any practical significance to a user of the forecasts. When looking at Graph 7.2, things turn a bit more complicated due to the number of lines. The line that deviate the most from the 45 degree line is IT which we can visually identify as being different from the 45 degree line. The sector that differs the most from the 45 degree line is IT followed by Telecom.



It is easy to see from the graph that the eight lines are not the same. It is also easy to see that most Maximum values for the regression lines are different from the 45 degree line value. However, the sample is not clustered around the maximum but rater somewhere in the middle as illustrated by table 6.1. From looking at the first two thirds of the lines we see that the differences are small and thus conclude that to an investor the practical difference is not significant. We thus conclude that all lines practically could be approximated with the 45 degree line.

7.5. Security Analysts' Earnings Forecasts for the Fourth Quarter are Less Accurate than the First Three (H5)

7.5.1. Percentage Error Analysis

The median APE of the Quarter 2 is 9.8% which is the highest error percentage of any quarter. However, as none of the differences between Quarter 4 and the other quarters are statistically significant at 1 or 5%, we can't say that the fourth quarter is less accurate.

Table 7.9 Median APE Quarters

	Median APE			Z
q4 vs q1	9,8%	>	8,9%	-0,51
q4 vs q2	9,8%	>	9,6%	-0,20
q4 vs q3	9,8%	>	9,0%	-0,48

*=sign. at 5%

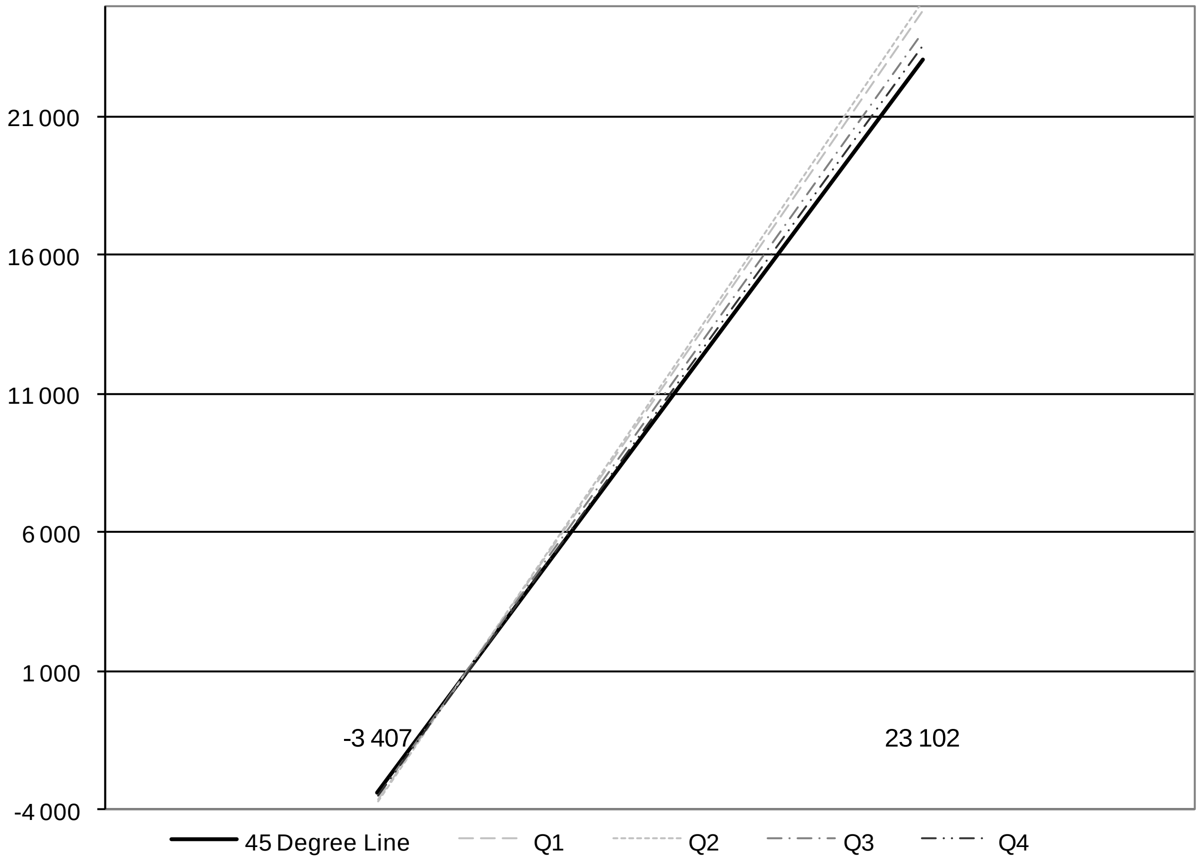
**=sign. at 1%

7.5.2. Regression Line Analysis

The r^2 values are fairly constant between all quarters ranging from 0.952 for Quarter 2 to 0.989 for Quarter 4 (see Table 6.5). This shows that the r^2values don't show any difference for Quarter 4 either, and we can thus conclude that there is no statistical difference between the quarters. The explanation could be either that SE's are fairly good at predicting the accounting effects in the fourth quarter, or that the accounting effects are not as severe as we assumed.

All quarters but Quarter 4 are statistically different from the 45 degree line at both 5% and 1% significance level. After a visual comparison of the four regression lines to the 45 degree line, we conclude that the deviations between the regression lines and the 45 degree line are of no practical significance to an investor. Thus, we can't say that Security Analysts' earnings forecasts for Quarter 4 are less accurate than for the other quarters.

Graph 7.3 Regression Line - Quarters



7.6. Security Analysts tend to overestimate results in a bear market and underestimate results in a bull market (H6)

7.6.1 Percentage Error Analysis

Both Bull and Bear tend to underestimate earnings with 58.8% and 58.2% of the observations respectively. This is significant at 5% for Bull and 1% for Bear. As both market trends underestimate earnings, we have to reject our hypotheses.

Table 7.10 RPE Quantity Market Trend

Market trend	Observ.	Underestimations			Z	
Bear	170	58,8%	>	50,0%	1,5	
Bull	884	58,2%	>	50,0%	4,9	**

*=sign. at 5%
**=sign. at 1%

The median RPE for Bear (3.23) is somewhat higher than for Bull (2.59), but these differences are not statistically significant at either 1% or 5%. We thus draw the conclusion that the underestimation pattern is not dependent on the market trend, but is instead a general trend. This shows that SE's are not reverting to mean, but are underestimating in both market trends.

Table 7.11 Median RPE Market Trend

Market trend	Median RPE	Z
Bear vs. Bull	3,2%>2,6%	-1,51

*=sign. at 5%
**=sign. at 1%

8. Conclusion

In this section we will make our concluding remarks in relation to the purpose we set up. For the first part of our purpose, we start off by concluding that Security Analysts produce significantly more accurate earnings forecasts than Times-Series Models. For an investor, this extra value added by Security Analysts is of such magnitude that the use of time-series models can't be justified. Despite Security Analysts' forecasting superiority over Times-Series Models, we also conclude that their forecasts still are not perfect.

Next, Security Analysts tend to underestimate results, which support the statement above that their forecasts are not completely accurate. We further conclude that this is a shortcoming from the Security Analysts' part, as this pattern is something they should be able to detect and correct. The Times-Series Models also generally underestimate results. The underestimation problem stems from an inability to incorporate the growth in earnings.

As for the second part of our purpose, we start by drawing the conclusion that Security Analysts' forecasting accuracy do vary with different circumstances. Of the four circumstances looked at in this theses, we found that the sector belonging affected the forecasting accuracy. We did however not see any considerable difference in forecasting accuracy between company size, fiscal quarter or market trend.

However, when analyzing whether or not circumstances affect the forecast accuracy, we realized that the answer to that question often depends on how you define and measure forecast accuracy. This could be highlighted when looking at our first circumstance, size. On one hand, smaller companies tend to make larger percentage errors, but this is rather a function of a measurement inaccuracy to measure performance, than of the size. Thus, we draw the conclusion that when analyzing forecasts for different circumstances, it is vital to succeed in separating the effect from the circumstance specific features from those coming out of other sources.

8.1 Reliability and Validity

In terms of reliability, it is our firm belief that the conclusions reached in this thesis are reliable and that the same results would be reached if the study was replicated, given the same sample and delimitations.

Even though some of the conclusions drawn are not in line with previous research in the area, we still find them robust since they are due to differences in the method of evaluating forecast accuracy. In most previous research, the percentage error has been the standard evaluation tool. We have added the regression analysis as a complement in order to remedy the measurement problems caused by small earnings when using the percentage error. Our results from the percentage error analysis are, however, mainly in line with the finding of previous research. Hence, we believe that our study has a high degree of reliability.

Regarding validity, we have used similar method and measurement as previous similar studies. In addition, we have improved the validity of this study since we have adjusted for earlier recognized measurement problems (measuring small earnings) by applying a regression analysis. We have also used a different definition of earnings that does not have the problems associated with using earnings per share, which has been used frequently in previous studies. We are, however, aware of that we have not taken into account that companies have been added or removed during the period. Since our sample is large and the number of companies added or deducted small, we don't see this as a problem. In addition, in the comparison between Security Analysts and the Time-Series Models, the exact same observations are used.

8.2 Suggestions for Further Research

We believe that we have reached some very interesting results in this paper mainly due to the novel analytical tools used. Since the main research on earnings forecasting have been conducted using the percentage error measurement approach, rather than the Regression Analysis approach, it would be interesting to reexamine some older studies and determine if the results have been biased by the size of earnings or make a replicating study on another sample. This is of extra interest, as we generally come to the same conclusion as previous studies when looking at percentage error.

We have looked at four different circumstances. One logic step would be to locate more circumstances and construct a multiple regression analysis in order to see the contribution on the forecast error from each circumstance. One example would be to look at whether the number of analysts making up the consensus is affecting the forecasting accuracy.

It would also be of interest to look at multivariate Time-Series Models, in order to see if they perform better than the rather simple Martingale and Sub Martingale models used in this thesis. Another interesting comparison would be to construct a model based on accounting figures and historical ratios.

As firms are generally followed by specific Security Analysts we believe that it would be interesting to compare the sectors across countries. There is also a possibility to incorporate the different circumstances in a multiple regression line and test the new models against a new sample.

9 Reference

Literature

Albrecht,W.S. Lockabill, L.L. & MCKeown, J.C. (1977)”The Time-Series Properties Anual Earnings” *Journal of Accounting Research* pp. 226-244

Ashton, A.H. Aschton, R.H. (1986) “Aggregative Subjective Forecasts, Some Empirical Results” *Management Science* pp.1499-1508

Ball, R. Brown, P. (1968) “An Empirical Evaluation of Accounting Income Numbers” *Journal of Accounting Research* 159-178

Ball, R. Watts, R (1972) “Some Time Series Properties of Accounting Income” *Journal of Finance* pp.663-682

Bathke, A.W. Jr. Lorek, K.S. Willinger, G.L. (1989) “Firm-Size and the Predictive Abillity of Quarterly Earnings Data” *The Accounting Review* pp.49-68

Brown, L.D. Hagerman, R.L. Griffin, P.A. Zmijewski, M.E.(1987) ” Security Analysts Superiority Relative to Univariate Time-Series Models in Forecasting Quarterly Earnings” *Journal of Accounting and* Economics pp.61-87

Brown, L.D. Richardson, G.D. Schweger, S.J. (1987) “An Information Interpretation of Financial Analyst Superiority of Forecasting Earnings” *Journal of Accounting Research* pp.49-67

Brown, L.D. Rozeff, M.S. (1978) “The Superiority of Analyst Forecasts as Measurement of Expectations: Evidence from Earnings” *The Journal of Finance* pp.1-16

Collins, W.A. Hopwood, W.S. (1980)” A Multivariate Analysis of Annual Earnings Forecasts Generated form Quarterly Forecasts of Financial Analysts and Univariate Time-Series Models” *Journal of Accounting Research* pp. 390-406

Conroy, R. Harris, R. (1987) “Consensus Forecasts of Corporate Earnings; Analysts’ Forecasts and Time Series Methods” *Management Science* pp.725-738

Crichfield,T. Dyckman, T. Lakonishok, J. (1978)“An Evaluation of Security Analysts’ Forecasts” *The Accounting Review* pp.651-668

Edlund, P-O. (1997) ”SPSS för Windows 95Multipel regressionsanalys Version 7,5” *Studentliteratur* Lund Swe

Elton, W.J. Gruber, M.J. Gultekin, M.N. (1984) “Professional Expectations Accuracy and Diagnosis of Errors “ *Journal of Financial and Quantitative Analysis* pp.351-363

Finn, F.J. Whittred, G.P. (1982) “On The Use of Expectations of Earnings per Share as Experimental Benchmark” *Economic Record* pp.169-173

Foster, G. (1986) “Financial Statement Analysis 2nd Edition” *Prentice Hall* Engelwood Cliffs NJ USA

Foster, G. (1977) “Quarterly Accounting Data: Time-Series Properties and Predicative-Ability Results” *Accounting Review* pp.1-21

Gujarati, D.N. (2003) “Basic Econometrics 4th Edition” *McGraw-Hill* 1221 Avenue of the Americas NY USA

McEnally, R.W. (1971) “An Investigation of the Extrapolative Determinants of Shor-Run Earnings Expectations” *The Journal of Quantitiative Analysis* pp.687-706

Newbold, P. Carlson, W.L. Thorne, B. (2003) “Statistics for Business and Economics 5th Edition” *Pearson Education Inc.* Upper Saddle River NJ USA

Ruland, W. (1972) “On the Choice of Simple Extrapolative Model Forecasts of Annual Earnings” *Financial Management* pp.30-37

Watts, R. (1976) “Comments on “On the Informational content of Dividends” *Journal of Business* pp.81-85

Watts, R.L. Leftwich, R.W. (1977) “The Time Series of Annual Accounting Earnings” *Journal of Accounting Research* pp.253-271

Online Sources

OMX Nordic Exchange
http://omxnordicexchange.com/forbolagochemittenter/produkterochtjanster/Sektorindelning May 4th 2008

College of Saint Benedict|Saint Johns University Physics department [intenet]
http://www.physics.csbsju.edu/stats/KS-test.html May 4th 2007

Financial Databases

SME Direct

Appendix

APPENDIX 1

A1.1 Pros nad Cons of Security Analysts and Time-Series Models

Security analyst approach to forecasting

Pros

1. Ability to incorporate nformation from many sources.
2. Ability to adjust to structural change immediately.
3. Ability to update continually as new information becomes avilable.

Cons

1. High setup cost and high ongoing cost to monitor numerous variales, make company visits, and so on.
2. Heavy dependence on the skills of a single individual.
3. Analyst may have an incentive not to provide an unbiased forecast (e.g. due to pressure to conform to consensus forecasts).
4. Analyst may be manipulated byb company officials (at least in the short run)

Univariate time-series model approach to forecasting

Pros

1. Ability to detect and exploit systematic pattern in the past series.
2. Relatively low degree of subjectivity in he forecasting (especially given the availability od computer algorithms to identify and estimate models).
3. Low cost and ease of updating.
4. Ability to compute confidence intervals around the forecasts.

Cons

1. Limited number fo information avalable for newly formed firms, firms with structurall change and so on.
2. Financial statement date may not satisfy distributional assumptions on time-series model used.
3. Inability to update forecast between successive interim or annual earnings releases.
4. Difficulty of ommunicating approach to clients (especially the statistical methodology used in identifying and estimating univariate models).

Appendix 2

A2.1 Kolmogorov-Smirnov Test

In order to test whether or not the sample may be assumed to be normally distributed we utilize a goodness of fit test. We have chosen the one sample Kolmogorov-Smirnov test (K-S test) as it is a simple way of detecting any differences in distribution compared to the normal distribution. There are several other models however they are far more cumbersome and have shown to be extremely sensitive to any differences between the sample distribution and the null hypothesis distribution by comparing a cumulative fraction plot of the sample with the control group. The use of fractions makes the K-S test insensitive to scale differences i.e. log. One example of alternative methods is the Anderson-Darling test which has shown to reject any imperfections with samples over 25 observations. As we have 836 observations this would not be an appropriate test. A problem with the K-S test is its ability to test not only the distribution but also the location of the distribution we however assume that this is not a cause for bias in the sample.

H0= $X \sim N(\mu, ð^2)$

H1=Sample doesn’t fallow the normal distribution

The null hypothesis is rejected if Z is above the critical value (1,96 for 5% and 2,58 for 1% significance level).[78]

A2.2 Outliers/ Extreme Values

As described earlier one of the main problems when dealing with ratios i.e. percentage error is that errors near zero has a tendency to create large percentage deviations; And thus cause a bias when looking at the mean of the sample. As the cause of outliers is not a simple function of erroneous observations but may be real observations that actually are as abnormal as indicated it is difficult to determine what a true outlier is.

This causes the problem of not only detecting but also of determining if observations are true outlier or extreme value. If a large percentage of the observations are identified as potential outliers and it is difficult to determine which are true outliers it may be easier to use the median; which is not affected to the same degree as the mean.

[78] College of Saint Benedict|Saint Johns University Physics Department [Internet] <http://www.physics.csbsju.edu/stats/KS-test.html> May 4th 2007

We have chosen to define outliers and extreme values as values beyond 1,5 and 3 times the inter-quartile value.

Outliers	$\geq$	[1,5*(q3-q1)+q3]
Extreme values	$\geq$	[3*(q3-q1)+q3]

q1=Quartile one
q3=Quartile three

A2.3 F-test

While the student t-test illustrated the degree of certainly of the individual variables (Slope and Y-intercept) the F-test compares both simultaneously. Thus it tests the degree of certainly we can rely on observed difference between the observed curve and the 45 degree slope.

$$F=\frac{\frac{RSS_r-RSS_{ur}}{m}}{\frac{RSS_{ur}}{n-k}}$$

RSS_r=	Residual Sum Squared restricted
RSS_{ur}=	Residual Sum Squared unrestricted
m=	number of linear restrictions
n=	number of observations
k=	number of variables

RSS_r is defined as the sum of the squared vertical deviations of each individual outcome and the regression line.
RSS_{ur} is defined as the sum of the squared vertical deviations between the individual forecast and outcome.

Hypothesis

H0; $\beta_0=0$

$\beta_1=1$

H1; $\beta_0\neq 0$or $\beta_1\neq 1$

We have included the highest level at which the null hypothesis may be rejected. Hence if the F-test has a value of 4,0 we can reject the null hypothesis at 5% significance level. This implies that given that the test is repeated 20 times we will accept the null hypothesis once even though it should have been rejected.

Decision rule $F\geq F_{m,n-k,\,\alpha}$ reject Ho with degrees of freedom m and n-k. [79]

A2.4 Practical Versus Statistical Significance

Statistical significance refers to the degree of certainly the conclusions provided by the tests may be relied upon. A conclusion may only be assumed correct if the statistical significance

[79] Edlund (1997) p.163

accepts the hypothesis. This is done by comparing the degree of certainly to a distribution. It may be regarded as significant if there is a 95 or 99% certainly of the results. However statistical significance does not reflect the relevance of the test in practice. Thus one often discusses practical significance. Practical significance may not be tested according to rigid formulas but is rather a judgment process. To illustrate the difference an example follows: A forecasting model gives a 5% statistically significantly different forecast compared to the outcome, in the amount of 50 000SEK. However if the forecasted outcome is ten billion it is unlikely that the difference of 50 000SEK would make any difference, thus there is a difference that can be proven statistically however it has no implication on the application of the uses of the forecast. The statistical significance of both t- and F-tests don't say whether a line is a good approximation or not, but only if you can or cannot reject that the two lines are identical. This means that if you have an extremely large sample, the tests may reject even a very small difference (as in the above mentioned example). The way the hypotheses is formulized it also means that a sample well gathered around the regression line may be rejected due to a very small difference, while a scattered sample may not be rejected as it is harder to exclude the possibility that it isn't the same line. This means that a rejected line automatically isn't worse than an accepted line (unless the r^2 values are equally good).[80]

A2.5 Discussion Regarding Alternatives to t-test.

A question not often discussed in papers dealing with earning forecasst evaluation is the inherent problem with using rations i.e. EPS and the statistical testing of the deviations between models. Lawrence, Brown& Rozeff published a paper in 1978that deales with the misuse of paired t-tests as a statistical evaluation of the errors. The main argument lie in tha the error is not drawn from the same population as each firm has a different EPS level and thus a t-test may not be used to evaluate the difference in means. Nor would the paired t-test be used when dealing with error rations as the effect from outliers can be assumed to be severe. This is due to the effect of observations around zero and the percentage effect of such values. We therefore must either assume a normal distribution or compare the data using a test that is insensitive to outliers and error definition.

A test that satisfies the insensitivity to the error definition problems and problems caused by ratios one such test is the non parametric Wilcoxon Signed Rank test.[81]

[80] Gujarati (2003) pp.138-139
[81] Brown & Rozeff (1978) pp.1-16

A2.6 Wilcoxon Signed Rank test

In order test if the median between to samples are different one may use the Wilcoxon signed Rank test not to be confused with the Wilcoxon rank-sum test, which is the same as the Mann-Whitney U test described in section 5.1.2.10. The Wilcoxon Signed Rank test is a used to test if two samples have the same median as each other. Basically it is the equivalent median test to the paired t-test which test the mean of two samples.

The test is carried out in by pairing the samples and calculating the difference between the pairs. Each pair with the difference of zero is excluded and the remaining pairs are ranked by absolute difference. If several pairs have the same difference they are assigned the average rank i.e. if three pairs are awarded rank 17, 18, 19 they all get 18.The rang for the positive differences are then added together to form T^+ and the same is done for all the negative rang's to form T^-.

H0; $Median_1$ equals $Median_2$ ($M_1=M_2$)

H1; $M_1 \neq M_2$

we sett T= lowest value of T^- and T^+. Since we have more than 20 pairs we can approximately use a normal distribution.

The test Z value is calculated as shown below

$$Z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)2n+1)}{24}}}$$

T= WSR rank value

n= Number of observations

The null hypothesis is rejected if z is above the critical value (1,96 for 5% and 2,58 for 1% significance level).[82]

A2.7 Binomial test

The binomial test allows us to statistically test the significance of deviation from an expected random outcome from a two outcome test. When testing a two outcome sample we assume

[82] Newbold, Carlson & Thorne (2003) pp.539-541

that the outcome should be 50% in each outcome.. (p=0.5). Assuming more than 10 observations in each outcome group we can approximate the binomial distribution with a normal distribution.

H0; p=0.5

H1; p≠0.5

The test Z value is calculated as shown below

:

$$Z = \frac{\frac{X}{n} - P}{\sqrt{\frac{pq}{n}}}$$

X= Number of observations in group one
n= Total number of observations
p= Fraktion expected in group one
q= Fraktion expected in group two

The null hypothesis is rejected if z is above the critical value (1,96 for 5% and 2,58 for 1% significance level).[83]

A2.8 Mann-Whitney U test

The Mann-Whitney U test is used in order to test if two independent population come from the same distribution which often is interpreted as whether or not they have the same median. The test relies upon that all the observations are ordinal or continuously measurements and that the two samples are independent. If the number of observations for each population is over ten we can approximately use the normal distribution as the basis for this test.

H0; ($M_1 \leq M_2$)

H1; $M_1 > M_2$

Every observation is assigned a rang, with the lowest error receiving one and so on and so forth disregarding which population the observation belongs to. If several have the same magnitude then they share the same average value. T is defined as the sum of all rang values awarded to population one.

[83] Newbold, Carlson & Thorne (2003) pp.147-153

The test Z value is calculated as shown below

$$Z = \frac{\frac{n1(n1+1)+n1n2}{2} - T}{\sqrt{\frac{n1n2(n1+n2+1)}{12}}}$$

n= number of observations
T= M-W rank value

The null hypothesis is rejected if z is above the critical value (1,96 for 5% and 2,58 for 1% significance level)[84]

[84] Newbold, Carlson & Thorne (2003) pp.543-544

Appendix 3

Table A3.1 Company sample

Firm	Sector	Size	Currency		Firm	Sector	Size	Currency	
ABB	Industrials	Large	USD		MTG	Consumer	Mid	SEK	
Alfa Laval	Industrials	Large	SEK		Munters	Industrials	Small	SEK	
Assa Abloy	Industrials	Large	SEK		NCC	Industrials	Mid	SEK	
Astra Zeneca	Healthcare	Large	USD		Nobel Biacare	Healthcare	Large	EUR	
Atlas Copco	Industrials	Large	SEK		Nobia	Consumer	Small	SEK	
Autoliv	Consumer	Mid	USD		Nokia	Information Technology	Large	EUR	
Axfood	Consumer	Mid	SEK		Nordea	Financials	Large	EUR	
Axis	Information Technology	Small	SEK		OMX	Financials	Mid	SEK	
Billerud	Materials	Small	SEK		Orc Software	Information Technology	Small	SEK	
Boliden	Materials	Mid	SEK		Oriflame	Consumer	Mid	EUR	
Cardo	Industrials	Small	SEK		Q-Med	Healthcare	Small	SEK	
Carnegie	Financials	Small	SEK		RnB	Consumer	Small	SEK	
Clas Ohlson	Consumer	Small	SEK		Saab	Industrials	Mid	SEK	
Electrolux	Consumer	Large	SEK		Sandvik	Industrials	Large	SEK	
Elekta	Healthcare	Small	SEK		SAS	Industrials	Small	SEK	
Eniro	Consumer	Small	SEK		SCA	Materials	Large	SEK	
Ericsson	Information Technology	Large	SEK		Scania	Industrials	Large	SEK	
Getinge	Healthcare	Mid	SEK		SEB	Financials	Large	SEK	
Gunnebo	Industrials	Small	SEK		Securitas	Industrials	Mid	SEK	
H&M	Consumer	Large	SEK		Securitas Direkt	Consumer	Small	SEK	
Haldex	Industrials	Small	SEK		Securitas Systems	Industrials	Small	SEK	
Hexagon	Industrials	Large	SEK		SHB	Financials	Large	SEK	
HiQ	Information Technology	Small	SEK		Skanska	Industrials	Large	SEK	
Holmen	Materials	Mid	SEK		SKF	Industrials	Large	SEK	
Husqvarna	Consumer	Mid	SEK		SSAB	Materials	Large	SEK	
Höganäs	Materials	Small	SEK		Stora Enso	Materials	Large	EUR	
IBS	Information Technology	Small	SEK		Swedbank	Financials	Large	SEK	
Intrum Justitia	Financials	Small	SEK	*	Swedish Match	Consumer	Large	SEK	*
JM	Finacials	Mid	SEK		Tele2	Telecommunications	Large	SEK	
Kappahl	Consumer	Small	SEK		Teleca	Information Technology	Small	SEK	
Kaupthing	Financials	Large	ISK		Telelogic	Information Technology	Small	SEK	
Lindab	Industrials	Small	SEK		TeliaSonera	Telecommunications	Large	SEK	
Lundin Mining	Materials	Mid	USD		TietorEnator	Information Technology	Small	EUR	
Lundin Petroleum	Materials	Mid	SEK	**	Tradedoubler	Information Technology	Small	SEK	**
Meda	Healthcare	Mid	SEK		Trelleborg	Industrials	Mid	SEK	
Metro	Cinsumer	Small	USD		Unibet	Consumer	Small	GBP	
Millicom	Telecommunications	Large	USD		Volvo	Industrials	Large	SEK	

*=Changes from Industrials
**=Changed from Energy

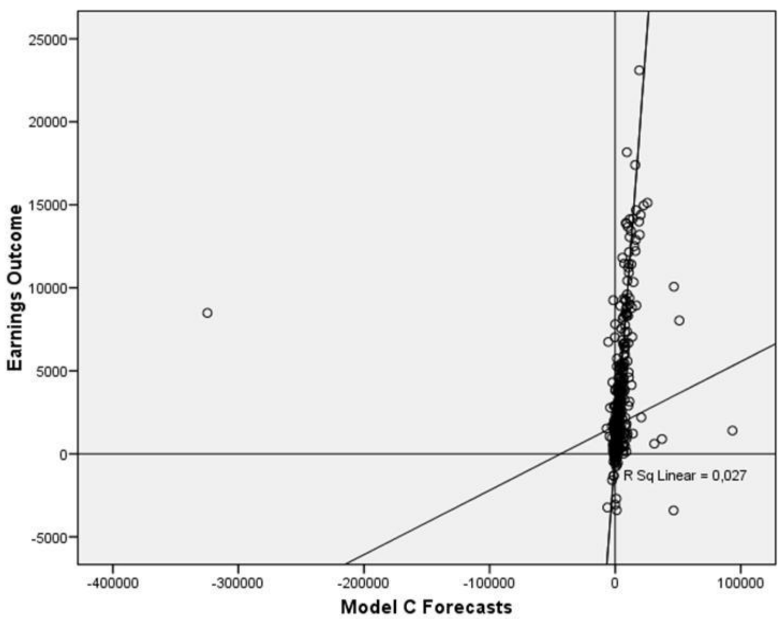
*=Changes from Industrials
**=Changed from Energy

Table A3.2 Exchange rate as of 1st of Januari

Currency	2001	2002	2003	2004	2005	2006	2007
EUR	9,31	9,12	9,06	9,02	9,41	9,04	9,44
GBP	..	..	..	12,73	13,68	13,40	12,86
ISK	..	..	..	..	..	0,10	0,10
USD	10,46	8,70	7,26	6,62	7,95	6,84	6,47

..= Not Avalible

Graph A3.1 Model C Scattergram



Appendix 4

Table a4,1 K-S test if APE/RPE are Normaly distributed

	SA	Model A	Model B	Model C
RPE (Z-values)	10,2 **	9,79 **	9,5 **	13,18 **
APE (Z-values)	11,71 **	11,36 **	11,32 **	13,85 **

*=sign. at 5%

**=sign. at 1%

K-S test of APE differences are Normaly Distributed

	SA-Mod A	SA-Mod. B	SA-Mod. C	Mod A-Mod B	Mod A-Mod C	Mod B-Mod C
Difference bet	10,17 **	10,21 **	13,44 **	10,18 **	13,34 **	13,21 **

Absolute error (Z-values)

*=sign. at 5%

**=sign. at 1%

Table A4.3 Regression for all models Size

Firm size	Forecast mod	Slope	R2 Value
Large	Cons. Mean	1,049	0,97
Large	Consensus	1,050	0,97
Large	Model A	0,955	0,806
Large	Model B	1,013	0,884
Large	Model C	0,27	0,017
Large	Model D	0,893	0,806
Medium	Cons. Mean	0,911	0,853
Medium	Consensus	0,913	0,858
Medium	Model A	0,722	0,393
Medium	Model B	0,882	0,543
Medium	Model C	0,08	0,172
Medium	Model D	0,675	0,393
Small	Cons. Mean	0,969	0,897
Small	Consensus	0,972	0,907
Small	Model A	0,741	0,635
Small	Model B	0,986	0,22
Small	Model C	0,031	0,098