LUND UNIVERSITY SCHOOL OF ECONOMICS AND MANAGEMENT

Bachelor's thesis in Quantitative Finance

A Development of a Quantitative Stock Selection Model for Swedish Mid/Large Cap Stocks

Author: Stefan Hägnesten

f04sh@student.lth.se

Tutor: Hossein Asgharian

Discussant: Daniel Nilsmo

Date: 2009-02-05

Abstract

In this thesis an extensive development of a multi factor model in the effort to predict stock returns in the Swedish markets is undertaken. The aim is not only to create an easy to use and yet accurate model to assist in investment choices, but also to create a better understanding of which of the underlying fundamental and technical values explain future performance of large and mid cap stocks. Twelve explanatory variables are being analysed through a strictly quantitative perspective and finally put together to one ready-to-use price predicting model based on six fundamental and technical values.

Acknowledgement

I would first and foremost like to thank Christer Swaretz at *United Brokers* brokerage firm for giving me the opportunity to collaborate with them with the ambition of developing a quantitative stock selection model. It has been a very stimulating project and I am deeply greatful. I would also like to thank Mathias Andersson for helping me with the financial background issues through out the project. Last, but not least, I would like to thank my tutor Hossein Asgharian at Lunds University for providing me with necessary technical information about the mathematics and econometrics.

Contents

Tables

1 Introduction

The stock markets around the world have historically all been characterized by periods with great volatility and nothing is pointing at a decrease. We have endured huge financial collapses, which all, starting with the Great Depression of 1929 to the credit crunch of today, have made impact on the global economy as a whole and everyone has suffered from these more or less, in some way or another. Heavy evidence shows the financial markets are very hard to predict and an investor needs great skill to make money from it in the long run. Real returns on the Swedish stock market has historically been around 7 %, Bäckström U, (2001), calculating with an inflation of 2-3 %, but there are vast amounts of investors who has realized returns much higher than this and thereafter made a fortune from the markets. There are though probably even more who are on the opposite side and have defaulted. The stock markets put great pressure on how the investor should act. What differences in strategies are there between the fortune ones and the less fortuned ones? It's hard to say other than that they are many. And they are substantial. Investment strategies can be of styles ranging from totally passive asset holdings, where the buy and hold rule is followed, to algorithmic trading where the investor uses advanced computers and executes all trades via automatic computer programs to increase the number of trades and decrease reaction times to movements in the markets as much as possible, and all in between. The markets have an ability to adjust to the way people are investing and performance depends much on the skill and advancements of the individual investor. Some even argue that there is no need to put to much effort in investing strategies as investors do, since the markets are anyway to dynamic and are in some way dominated by the impact of improbable events to an higher degree than we in general believe, Taleb N. N. (2007) and (2004), which can make anyone with a "good" timing extremely rich or extremely poor. Active management to some level is though dominating the markets and the ambition with this project is to investigate the methods behind active management, stock selection and stock ranking. To realize a higher return than the historical one, the investor though needs, as mentioned above, skills. Active management that only reflect the investor's belief of out performing assets is a common strategy and the investor's ability to "master the markets" and separate the high return generating assets is of crucial importance. As the decision to choose an active management in front of a passive one relies on the belief that the investor knows something other investors do not and has an ability to choose outperforming stocks in a unique way, why settle with this strategy when the investor with the given skill can generate even higher returns by also separating underperforming stocks? These strategies are relaxing the long-only constraint and gives new opportunities to generate alpha, they are called active extension strategies. With a possibility to generate alpha both on the long leg and the short leg of a portfolio, an investor widens the investment universe. In a long only strategy, the maximum an investor can express his/hers negative view of an underperforming stock is underweighting it by its actual weight, but by relaxing this constraint an investor can underweight an underperforming stock more, the weight can become negative and he/she takes a short position in the stock. E.g. Armfelt C and Somos D. (2008) show that pools of such portfolios outperform long only portfolios with an equal alpha model. The purpose of this project is to construct a quantitative tool to assist an investor when deciding what his/hers views on stocks are, both positive and negative, and to separate these stocks from the others. The project is presented by *The United Brokers* brokerage firm and is supposed to be done on the basis of shorting 30 % of a portfolio and using these proceeds to go long in 130 % of the portfolio, a 130/30 active extension strategy. As argued above, the greatest complexity lies in choosing which 30% of the portfolio is believed to under perform and which 30% is believed to outperform and this is where the quantitative model comes in. With this project the edge between fundamental and quantitative analysis is cleared out in the effort of constructing the most accurate model possible.

The hypothesis is that stock returns one month ahead can be predicted with information available today and is therefore challenging any form of market efficiency. The company specific, fundamental variables consist of six fundamental ratios, four momentum variables, market capitalization and current price. A one month ahead predictive ability can be exposed on Swedish mid- and large capitalization stocks if the following five requirements are realized; the information set is large enough, the right factors chosen as well as the number of factors, the models are adjusted to the target market and the most appropriate quantitative methods are used. This hypothesis is the base of building a quantitative stock return-predicting model and is used in the project as follows; the universe for the model is the Swedish large and mid cap stocks and an extensive set of data is collected for the years 2000-2008 from Thomson-Reuters Knowledge database. The quantitative method used is linear regression with fixed effects in panel data with one month stock returns as a dependent variable and twelve fundamental, company specific values as independent variables. The quantitative models developed are put to test through several “out-of-sample” tests consisting of Swedish small cap stocks, Nordic mid and large cap stocks (excluding stocks traded on the Stockholm stock exchange) and Swedish mid/large cap stocks not used in the constructing of the model. A heavy analysis of the results and the performance of the models are done so that one model can be pointed out to be the final winner.

2 Earlier Research

2.1 Active Extension Investment Strategies and the value of Active Management

Johnson, Ericson and Srimurthy (2007) showed how an active equity strategy with a removal of the long only constraint adds value in the pursuit of a higher alpha. They also argue for the appropriate way of basing investment decisions on a quantitative rule. The study compared historically built 130/30 strategies to long only strategies, which showed a significant magnitude of the out performance of a quantitatively based 130/30 portfolio using six factors. A comparable research was made in a master thesis work by Almfelt and Somos (2008). They analyzed the performance of an active extension strategy, which out performed a long only strategy when based on the same alpha model for an 80 years long time horizon. The results concluded a 150/50 strategy generating the highest returns.

Another interesting article is Clarke R, de Silva R and Thorley S (2001) in which they divide the added value by actively managing portfolios into both the forecasting ability and the ability to take appropriate positions in securities that reflect those forecasts. They extend the fundamental law of active management, which gives the maximum expected value added to an actively managed portfolio, to containing also a transfer coefficient, being the correlation between the forecasted risk-adjusted returns and the risk weighted exposures of the assets in a portfolio, as well as an information ratio, being the expected excess return over a benchmark transformed into standard deviation space, and therefore deepens the dimension one needs to consider when choosing to actively manage portfolios. Clarke R, de Silva H, Sapra S and Horley S (2008) further increase the investors understanding of the influence of factors, such as stock correlation, on the size of short extensions in long-short strategies. When considering transaction costs and leverage costs active extension portfolios decrease in efficiency which is something Sorensen E, Lingjie M, Hua R and Qian E (2007) underlines, they though argue that the full relaxation should be limited to *some* relaxation, as in a 130/30 strategy. Transaction costs and stock loans are also considered by Martielli J. D. (2005) who in his educational article gives a very general understanding of the positive and negative aspects of an active extension portfolio.

Over all, in literature, the views of relieving the long-only constraints are positive, but must fundamentally be combined with the skill of choosing which stocks to underweight and which to overweight.

2.2 Multi Factor Models and Model Building

The most extensive research material I found covering the subject is made by Levy, Kenneth N. (2000, part 1). He underlines his belief in the market being defined as a complex system, see below in *Theory 3.2* for definition, and that one can add value to investment decisions from quantitative multifactor models such as this one, though one can not repress the seemingly random factor in market movements. He gives numerous examples of earlier research only examining two or three factors at once and argues for having more explanatory variables to decrease misleading results. He examines 18 separate factors, which are to explain interrelationships of equity return regularities. He presents both quantitative and economic arguments for each factor.

Many studies cover what regularities explain stock returns the best, as Jones C. P (2008) who determines the most significant factors being earnings growth, P/E ratio and the dividend yield, de Bondt G. J. (2008) pushing on the influence of fundamental values in the long run and factors as momentum and seasonality differences in the short run. Alford A, Jones R. C. Lim T, Litterman B (2004) argues strongly about the fundamental values driving alpha and the resulting model contains six variables being valuation, profitability, earnings, management impact, momentum and analyst sentiment.

Apart from the above mentioned articles, there are a few which should be mentioned regarding how to actually build models like the one developed in this thesis. Among these are *Quantitative Equity Analysis* by Team-Sys which very generally goes through the key elements of such a model, Burmeister E, Roll R, Ross S. A., Elton E. J., Gruber M. J., Grinold R., Kahn R. N (1994) who shows the strength of multi factor models by studying portfolios consisting of stocks from the S&P 500. It also presents the subject starting with APT and CAPM in a straightforward way. Brush J. S. (2006) and Nicholas, J G., (2000) focused more on dividing explanatory factors into value and growth factors and showed how they both influence stock returns in a similar matter and also exposes a common fallacy being that value would beat growth stocks. Growth factors are measures of a company's growth potential as can be reflected in for example future earnings or changes in investor expectations while value factors assess a company's current capital value, historical earnings stream or futures earnings prospects and then determine how much investors have to pay to attain this value, see Nicholas, J G., (2000, p208-210) for further details on selecting stocks by factors.

There are though some controversy of beating the market following simple quantitative rules as Cooper, Gutierrez and Marcum (2001) underlines. They use a recursive out-of-sample method to find that the possibility to beat the market is exaggerated in literature, when based on three factors company specific factors. Ferson W. E. Sarkissian S, Simin T. (1998) points out a common factor probably inducing bias in studies like this one; unbalanced data sets with more stocks than time periods.

Rapach E. D, Wohar E. M. (2004) find that there is not such a big difference between results from out of sample and in sample tests of models which describe stock return predictability in the context of data mining, to extract hidden information form data sets. This puts away pressure on having only out of sample tests, which can be to good help in this project.

3 Theory

A quantitative equity manager has two main tasks. The first one being to detect mispriced securities and the second one to combine a set of these into a portfolio while preserving superior returns and not incurring undue risk. When a security is mispriced it has the potential to provide superior returns when its price corrects over time. The price of a security, which is over priced, will according to the theory of equilibrium markets decrease and one that is over priced will increase. This change in price will continue until they both have reached equilibrium prices. To benefit from mispricing one needs to be able to exploit this phenomena and if this is possible is still a very discussed topic regarding if the markets are efficient or not and if so, to what degree. To continue, the meaning of an efficient market needs to be cleared out.

3.1 Efficiency in Markets

Efficient markets are all defined by the efficient market hypothesis, EMH, see Fama E. (1970). This hypothesis asserts that all financial markets are efficient in an informational degree. This means that prices of all securities are fully reflecting all known information and implies that there is no possibility to benefit from using already known information about any security. The hypothesis is classifying a market to belong to one of the following forms:

Weak form efficient: All historical prices of a security are reflected in today's price and no excess returns are possible based on historical security prices.
Semi-strong form efficient: all public information is reflected in the price of a security. No one with only publicly available information can systematically exploit excess returns.
Strong form efficient: All public and insider information is reflected in the price of a security and there is no possibility, whether one has access to insider information or public information, to benefit from excess return.

This hypothesis is though constantly being put under a magnifying eye and criticized in modern theory. It is highly controversial and of course none of the authors behind the books, journals and studies lying ground for this study is a believer of the EMH and each contain different criticisms of the market not being efficient. For example, one of the more common problems with the EMH is the widely used fact that stocks with a low price to earnings ratio is out performing other stocks. Any market crash or financial bubble is mysterious from the perspective of an efficient market hypothesis believer. If the markets are not efficient, but yet not non-efficient either, since anyone then could invest according to certain pre set rules and get infinitely rich, what kind of system are the markets of?

3.2 The complex financial markets

Jacobs, B. I. & Levy, K. N. (2000, ch. 1 p. 25) gives a very scientifical classification of the financial markets. They do so using taxonomy from the sciences with three types of systems existing; ordered, complex and random.

"Ordered systems are simple and predictable, such as the neatly arranged lattice of carbon atoms in a diamond crystal. Similarly, Newton's Laws of Motion are a simple set of rules that accurately describe the movement of physical objects. At the other extreme, random systems are inherently unpredictable; an example is the random behaviour, or Brownian motion, of gas molecules. Complex systems fall somewhere between the domains of order and randomness. The field of molecular biology exemplifies complexity. The mysteries of DNA can be unravelled only with the aid of computational science. The human mind alone cannot cope with DNA's complexity, nor do simple theories suffice. The stock market, too, is a complex system."

From this classification, security pricing is not easily explained by simple rules nor totally random, a market has a structure of a very complex web connecting interrelated return effects. This view of the markets is the laying ground for most of the studies dealing with the pursuit for new ways of explaining and, in some ways naive attempts, to systemize security returns. During the 1970s the largest markets of them all, the US market, was believed by academicians to be efficient. The belief was that no one, except insiders, could beat the market and the market efficiency was of a semi strong form. Passive management was naturally born, as it did not make sense to have an active strategy. As said above, the EMH was soon starting to be more and more criticized and the amounts of studies disproving it increased vastly.

3.3 To Disentangle the Complex Web

When the market is defined as a complex system with security returns being structured as a complex web, what is left to do is to disentangle the web. This is where the focus of this thesis lies. There are two fundamental questions one needs to answer to achieve this;

1. What is the web built of? What regularities explain security returns?
2. How is the web built? How are the return regularities connected?

To answer question one, the literature examining return regularities in the stock market brings forward approximately 35 factors, which are the most used ones and are documented to have the strongest return predicting abilities, see chapter *10.1 Appendix*. These can of course be defined in many different ways but the ones shown in the appendix are the most common ones discussed. The first subgroup, company specific, are the most appropriate to use when one wants to predict stocks on an individual level since changes of these are less likely to be dependent on changes of the same for another company. Changes in the macro economic factors, how ever, most often imply changes in whole groups, sectors etc. The company specific factors brought up in literature can be grouped in the following four subgroups; growth, value, liquidity and technical. To invest in growth or value stocks are two different investment strategies. As Dow, C. G. (1998) writes, the debate on which strategy is the best has been going on for years and is more a question of what style of investment you have. When investing in value stocks, the investor looks for stocks that are mispriced. The common signs of stocks being so have a low/high price to earnings, price to book, price to sales and high dividend yields. Such an investor is regarded to not believe in any kind of market efficiency. On the other hand, an investor who is looking for growth stocks often believe in some form of market efficiency since instead of looking for mispriced stocks, he/she is looking for stocks with a high growth potential. Low price to earnings, price to book etc are indicators of this, as well as a high momentum and significant revisions from analysts. Momentums is simply a measure of a stock's trend strength, see Bernstein J. (2002, p 17), and is calculated by subtracting yesterdays stock price from today's. Rate of change can also be used using division instead of subtraction. When the momentum is negative, it is a sign of the stock being in a negative trend and vice versa for a positive momentum. There are basically three different kinds of signals a momentum indicator can provide with, leading, lagging and time current. For the true definitions of these ones, see Bernstein J, (2002), but in short, if the price momentum together with the price increases, the indicator shows a continuing positive trend and a price, which rises together with a momentum decrease tends to precede a top. Falling price with rising momentum tends to precede a bottom. There is though some controversy about what time lags give an indication of a trend and what give an indication for reversion. A stock being in a heavy uptrend can also bounce back and reverse towards some old mean, see also Nicholas, J G., (2000, p210). Another strong indicator which can be used is the above mentioned estimate revisions. When the information universe is as huge as it is an investor can hardly keep track with all changes occurring, but if one is in a situation where a large part of the analysts change their mind in any way, there is a clear sign of something happening on a company individual level. A high liquidity is also an indicator of an information flow hitting buyers/sellers. The subgroup with technical indicators are all indicators used in technical analysis and a common one used is the moving average which is an average of a small subgroup of the total data set and is continuously calculated through time. An example is a 20-day moving average, which is simply a rolling average of the latest 20 days. To answer the second question above an extremely used and straight forward technique is regression analysis.

3.4 Regression Analysis

When one wants to make statistical inference about a given data set, two steps needs to be considered.

1. What does the best fitted model look like?
2. How good is the best fitted model?

To answer these questions, regression analysis and further analysis of the models properties can be done according to the steps in chapters *3.4.1-3.4.3 Theory*.

3.4.1 The Model

Using regression analysis one can obtain estimates of parameters in a relationship. For example, the simplest regression analysis model is where one has two time series in which one of them, X, is said to be explanatory of the changes of the other, Y, in a linear fashion. This model is described mathematically as

$$Y_i = \beta_1 + \beta_2 X_i + e_i$$

To every regression model there are different assumptions made and the basic assumptions for the model above are discussed in Dougherty C. (2007, p68). This model can easily be extended and generalized to contain several explanatory variables with each variable, both independent and the dependent variable, being a function of the true variable.

$$Y_i' = f_0(Z_i) = \beta_1 + \sum_{j=1}^{K} \beta_{j+1} f_j(X_{ij}) + e_i$$

The most important assumption in these models is that the modeller is assuming that every value of $f_j(Y_i)$ has two components and is dependent on these in a linear fashion, the non random part, consisting of the constants, β_1, and the sum of the functions of the fixed quantities, $\beta_j f_j(X_{ij})$, and the last component being an error term which is stochastic. To estimate the j+1 number of parameters in the model one needs to solve the minimization problem formulated as the following. First define the residual as

$$e_i = Y_i - Y_i' = Y_i - (\beta_1 + \sum_{j=1}^{K} \beta_{j+1} f_j(X_{ij}) + e_i)$$

and solve

$$\min_{\beta_i} \sum_{i=1}^{k} e_i^{\,2} = \min_{\beta_i} (Y_i - (\beta_1 + \sum_{j=1}^{K} \beta_{j+1} f_j(X_{ij}) + e_i))^2 \qquad (1)$$

where the β_i:s are the model estimates. To estimate the parameters in this way is referred to as the ordinary least squares technique, OLS, and is by far the most popular one used in regression analysis. When the parameters are estimated an important measure of how good a model is can be calculated in a similar fashion as equation (1), taking the sum of the squared residuals. This of course needs to be small for a good model.

3.4.2 Panel Data with Fixed Effects Regression

Panel data are situations where the data set is containing several observations over time for a set of individuals, being for example a set of N stocks and their K number of properties, variables, observed during an T long time period, see chapter *4 Data*. The key issue to understand here is that the models described above cannot be used in the same fashion any more, see Arrelano M. (2003, p11-16), and needs to be adjusted to having a data set

$$\left\{(y_{i1},...,y_{iT},(x_{i1},...,x_{iT}))_1,...,(x_{i1},...,x_{iT}))_K, i = 1,...,N\right\}$$

and a model

$$Y'_{it} = f_0(Z_{it}) = \beta_1 + \sum_{j=1}^{K} \beta_{j+1} f_j(X_{ijt}) + \eta_i + e_{it}.$$

What is characteristic in this model is that the η_i:s are constant through time but varying between the i unities why it is called fixed effects regression. For more information about how to use this model, see chapter *5.1 Method*.

3.4.3 The Properties – How Good Is The Model?

To find out how accurate a panel data model, such as the above, is at describing the studied phenomena, there is an endless dimension of analysis, which can be considered. The most common ones are discussed further below.

Sum of Squared Residual - How big is actually the result from (1)? What is the average error? This measure is easiest used in comparison between models.
T test – How sure are we about the explanatory variables being explanatory at all? How is the explanatory power of one variable compared to another?
Adjusted R2 – How much of the change in the dependent variable does the explanatory variables describe?

As the reader can see below in chapter *5 Method* and *6 Results and Discussion*, there are more ways to analyse the data, but the three indicators above is a good starting point. For complete definitions on the statistical evaluators, see Dougherty C. (2007).

4 Data

As Bruner, R. P., (1998, p236-237) argues, garbage in, garbage out. The data quality issue is a very substantial one and needs to be taken into consideration deeply. The model only gets as good as the data you put in to it. A fact is that financial reporting is not an exact science why it is of crucial importance to put in data from one data source only. The data used in the model must be representative of the entire population and a good way to solve this is to use the model with the exact data source it is built upon. For this project all the data is collected from Thomson-Reuters Knowledge database. The largest possible universe of stocks is chosen which is a set of 80 mid and large cap stocks with a time horizon being eight years. As a general rule for achieving statistical significance and accuracy in the results the greater the universe, the better accuracy, and this universe is the greatest possible with the information sources available. The coverage of data is differing substantially between the different values and the different companies, but the least number of data for each variable for each company belonging to this universe is one. The common divisor of frequency for the different variables is quarterly updates, why this needed to be set as the standard for the analysis. The last closing price is the variable, which has the best coverage with only 14% missing data for the whole universe while price/cash flow is the variable with the least coverage with 45% missing data. Since price is available in daily frequency, the one-month returns are being defined as the price change from the end of every quarter until one month later. With these definitions the union between the sets of information is zero, which is necessary to keep a statistical accuracy. Differences in how company specific values are defined and measured could add a negative bias in the results if they are mixed together, but since all the information is collected from the same database with a guarantee of standardized definitions, this will not be the case for this project. Going through all factors listed in chapter *10 Appendix* and sorting out those, which are less covered since 2000 to the degree of being not usable, the following factors are left:

- **Price/Book**
- **Price/Sales**
- **Price/Earnings**
- **Dividend yield**
- **Price/cash flow**

- **Price -** (referred to as low price in the continuation, because of the low price effect)
- **Price Momentum, 1 month**
- **Price Momentum, 2 months**
- **Price Momentum, 3 months**
- **Return on equity Momentum, 1 year**
- **Market capitalization**
- **Return on equity**

These are also consistent with being the most used ones; see for example Jacobs, B. I. & Levy, K. N., (2000), Bruner, R. P., (1998) and Nicholas, J G., (2000, p208-210). B. I. & Levy, K. N., (2000) also provide an extensive list of literature in where the importance of these factors is discussed thoroughly. These ones have the strongest documented prediction ability and are for these two reasons chosen out and tested in this thesis. Thomson-Reuters are defining the company specific values as the following;

Price/Book: Historic Price/Book value, LI
Price/Sales: Historic Enterprise Value/Revenue, TTM.
Price/Earnings: Historic P/E Excluding Extraordinary Items, Avg. Diluted Shares Outstanding, TTM.
Dividend yield: Historic Dividend Yield, common stock primary issue, %, TTM.
Price/cash flow: Historic Price/Cash Flow per Share, Avg. Diluted Shares Outstanding, TTM.
Price: Last closing price, currency SEK.
Return on equity: Return on Avg. Common Equity, % (Income Available to Common Excluding Extraordinary Items), TTM
Market capitalization: Historic Market Capitalization, Total Shares Outstanding, FI

LI is for Last Interim, TTM is for Trailing Twelve Months and FI is for Fiscal Interim.

Notice that the price momentum variables effects do have a bit of controversy, some believe they can cause a more reversal movement of the price while some believe they help a movement of the price to continue in the same way, as if the stock is in a trend. The time, which the momentum effect is measured, is the most controversial one with a shorter time showing more of a momentum effect while longer times show more of a reversal movement.

The price momentum variables are defined in the same fashion as return with the price momentum since j months back for stock i at time t being;

$$pricemom = \log\left(\frac{p^{i}_{t}}{p^{i}_{t-j}}\right)$$

This is what is referred to as the logarithm of the rate of change, ROC, in Bernstein J, (2002, p19), and is used due to its normalizing definition for better comparison between stocks. Return on equity momentum for stock i at time t since one year back is defined as the difference between return on equity for stock i at time t subtracted with return on equity at one year prior;

$$ROEmom = roe^{i}{}_{t} - roe^{i}{}_{t-1year}$$

All the stocks were chosen after which ones had a market capitalization greater than or equal to 150 million Euros at the first of July 2008, in the end of this project, list in chapter *10.3 Appendix*. Out of these 80 companies, 36 were large cap stocks and the rest 44 were mid cap stocks at this date, though these numbers of course change during time and to make sure a non-significant part of these moved down in market capitalization and became low cap stocks or companies moving from large cap to mid cap each companies market capitalization was studied for the whole period. The definition of mid cap at the first of July 2008 was all stocks having a capitalization above 150 million euros (approx 1 425m Swedish Kronors 20080701) and large cap all stocks above 1 000 million euros (approx. 9500m Swedish Kronors 20080701). Since the market capitalization value for a company constantly changes with a change in price and shares outstanding, the stocks did not always count as large or mid cap. In the universe spanning over the eight years with a frequency of quarters, the total percentage of times any stock reaching outside and becoming a small cap stock was 19% which possibly can cause bias in the results. This bias could make the model loose precision for the target market, but widen its usable environment to include small cap stocks as well. This is though an unwanted effect. For more details on how the market capitalization changed within the chosen universe, see chapter *10.3 Appendix*)

5 Method

5.1 The definition of the Model

The return of stock i is defined as the log of the price at a time one month ahead in relation to the current price;

$$r_{i(t+1)} = \log(\frac{p_{i(t+1)}}{p_{it}})$$

Functions of the j number of fundamental and technical values for stock i at time t are treated as explanatory variables and the return of stock i at time t+1 as the explained variable in a linear dependence with a company specific intercept, alpha, and an error term u:

$$r_{i(t+1)} = \log(\frac{p_{i(t+1)}}{p_{it}}) = \beta_1 + \sum_{j=1}^{K} \beta_{i(j+1)} f_j(X_{itj}) + u_{it} \qquad (2)$$

This is what is referred, as a linear forecasting regression model and the techniques from OLS regression theory is also the most common way to calculate the weights in this type of stock selection model. When a large amount of different unities are observed over more than one time point, the data structure is a panel data set and in this project panel data with fixed effects regression is used. The theory behind fixed effects is that the explained variable is explained by explanatory variables in the same fashion as equation (2), but every company does have a company specific factor, which adds up a term, which is fixed over time but differs between unities. This implies that for every company, the linear relationship looks like equation (3).

$$r_{i(t+1)} = \log(\frac{p_{i(t+1)}}{p_{it}}) = \beta_1 + \sum_{j=1}^{K} \beta_{i(j+1)} f_j(X_{tij}) + \eta_i + u_{it} \qquad (3)$$

η_i are the fixed effects. To be able to do an efficient analysis of such a set of panel data, the company specific effect must be eliminated so that every unit can be directly compared to any other unit in the set. The factors are thereafter affecting the companies in the same manner with common coefficients. To reduce this company specific fixed effect an entity demeaned OLS algorithm can be used. There are two steps; the first being to calculate the entity specific average for each variable including the returns and then subtract all these from each variable respectively. Equation (3) then takes the shape according to equation (4).

$$r_{t+1} = \log(\frac{p_{t+1}}{p_t}) = \sum_{j=1}^{K} \beta_{(j+1)} g_j(X_{tj}) + u_t \qquad (4)$$

Here the company specific effect is eliminated and the linear relationship is equal for each company, notice the eliminated i for each unit and the intercept. The dependent variable in equation (4) is each return, for each company i and each time t, subtracted with the average return over all companies and times. The independent variables are transformed variables according to the same method and the regression is now practically performed without an intercept, alpha equal to zero. When the coefficients are calculated, the second step is to reconstruct the model so it can be used for any chosen company which is regarded to lie in the same category as the companies in the panel, in our case any Swedish mid or large cap stock. The intercept needs to be reconstructed so equation (2) can be used again. To reconstruct the intercept, one assumption needs to be taken which is that the intercept is common for all companies and is the difference between the means earlier calculated for the returns and all the factor means multiplied with the now known coefficients;

$$\alpha = \frac{1}{N}\sum_{i=1}^{N}\overline{\alpha}_i = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\left(r_{it} - \sum_{j}^{K}\beta_{i(j+1)}\overline{g(X_{itj})}\right)$$

When the intercept and the coefficients are known, the final model, which can be used for forecasting stock returns, ends up like equation (5).

$$r_{t+1} = \alpha + \sum_{j=1}^{n} \beta_{(j+1)} f_j(X_{tj}) \qquad (5)$$

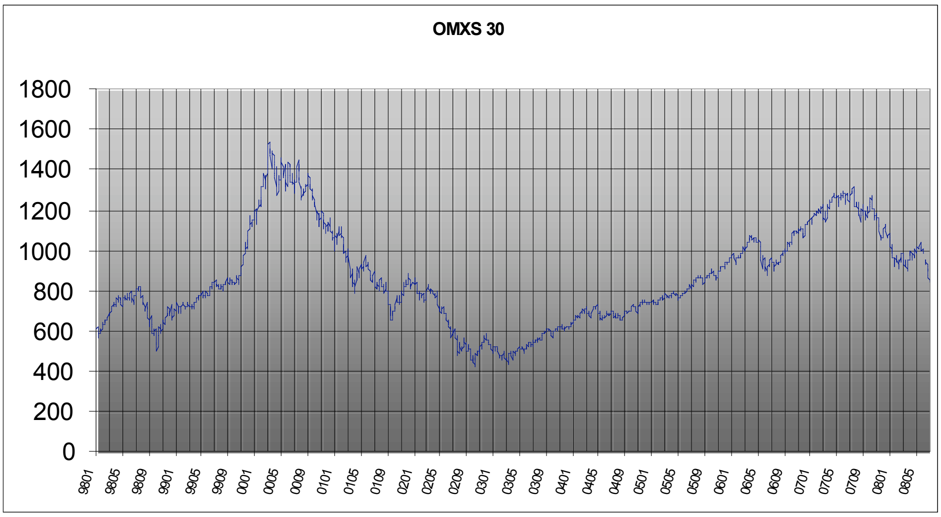
Notice that this linear relationship is now not specific for any stock and can be used in general for all stocks, which are regarded to belong to the target set. The company specific effect is eliminated and the return is only a linear relationship of j number of chosen factors.

5.2 The Use of the Model

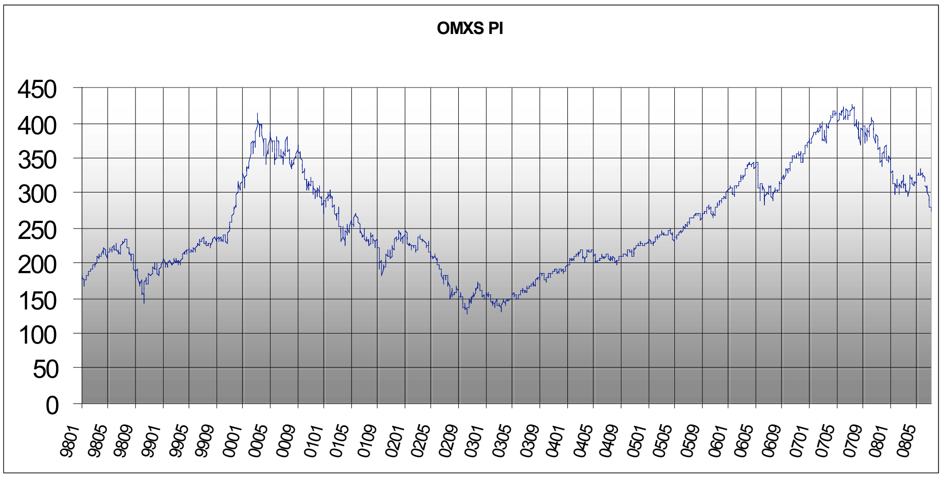
With a defined universe and a statistical toolbox available there are a couple of obstacles that needs to be taken care of before beginning. The following questions must be answered;

1. What different market situations does the time horizon include?
2. Can the universe be divided into subgroups with significantly different intra related behaviours, but common inter related behaviours?
3. How can missing data points be dealt with?
4. What number of factors and which factors do predict stock returns the best?

To answer the first question an index analysis is an appropriate way to go. To clarify how the Swedish stock market performed between years 2000-2008 we can observe two different indexes, the OMXS 30 and the OMX PI. Below are the two different indexes presented (taken from Thomson-Reuters).



Picture 1 price development of the OMX S30 index



Picture 2 price development of the OMX PI index

It is clear that they are highly correlated and show the same highs and lows. At first a very steep and volatile down period can be seen between 2000 until 2003 followed by a more consistent up period which ended in 2006. Between 2006 and 2007 the market had time to make a steep dip and then a quick rise followed by the last years financial tumult which caused the market to drop down quick again. Conclusively there are three major periods, down between 2000-2003, up between 2003-2007 and a down between 2007-today. In the continuation the ups and downs of the market is separated into five different sub horizons;

1. 2000-2008
2. 2003-2008
3. 2003-2006
4. 2006-2008
5. 2007-2008

While sub horizon 1 covers the whole time period, sub horizon 2 is partly defined to eliminate the number of down periods to as many as ups, but also with the benefit of reducing the percentage of data missing since the data coverage is poor the earlier the time observed. The third one is created to eliminate any downs and only regard a long, slow up going in the market while the fourth one is created to be able to catch a more volatile up and down. Last, the fifth one only contains a severe down period.

To answer question number two we need to take a closer look at the defined universe of stocks. Since the biggest issue is, as discussed in chapter *4 Data*, missing data, three different strategies for creating portfolios are chosen with the hope of not suffering from any kind of unbalance created by such missing data;

1. Create groups according to three different subgroups in themselves according to the data coverage of all the twelve variables. This results in one group containing the whole universe with all being covered at least at one out of the 33 time points (equals a poor 3% available data requirement), one with at least 16 out of 33 time points covered (equals at least 50% covered data) and last but not least a limit at 8 individual time points covered (equals at least 75% data available). They will be referred to as the 3%, 50% and 75% groups from now on.
2. Create groups according to which business sector they belong to according to Thomson-Reuters own RBSS Economic sector code. The universe is divided in 9 different sectors, see chapter *10.3.2 Appendix - RBSS Sectors*.
3. Create groups according to their market capitalization, most appropriate is one for medium cap stocks and one for large cap stocks.

When these three styles of dividing the universe are combined, there are 25 different settings for a continuation with quantitative analysis available;

2000-2008

1. 9 groups according to sector.
2. 2 groups according to market cap
3. 75% group
4. 50% group
5. 3% group

2003-2008

1. 9 groups according to sector.
2. 2 groups according to market cap
3. 75% group
4. 50% group
5. 3% group

2003-2006

1. 9 groups according to sector.
2. 2 groups according to market cap
3. 75% group
4. 50% group
5. 3% group

2006-2008

1. 9 groups according to sector.
2. 2 groups according to market cap
3. 75% group
4. 50% group
5. 3% group

2007-2008

1. 9 groups according to sector.
2. 2 groups according to market cap
3. 75% group
4. 50% group
5. 3% group

After the creation of the subgroups an extensive statistical examination is brought out on all the individual groups in two steps. The first step is to individually, factor by factor do a simple regression with returns to screen the predictability and statistical significance of each individually. In this step all the factors are examined both in original form, as defined above in chapter *4 Data*, but also transformed into a few different usual transformations to tailor make them individually and create the best fit possible to predict the return, see B. I. & Levy, K. N., (2000, p55-p59). The following transformations are examined; positive/negative effect, natural log, tenth log, second to eleventh root and second to ninth power. A one-sided hypothesis test is applied to test the significance of the factors, B. I. & Levy, K. N., (2000). The hypothesis is for all variables

$$H_0 : \beta_i = 0$$
$$H_1 : \beta_i > 0$$

This is why the factors that are supposed to have a negative effect on returns are transformed to a negative variable so that the alpha become positive and a one sided positive t-test can be used. The significance levels are chosen to 5 %, 1 %, 0,01 % and 0,001 %

The best transformations are thereafter used in a more complex multiple regression analysis. The twelve variables are regressed together with the returns and screened and controlled for statistical phenomenon's like co-linearity and relative ability of predictability and relative significance and how much information which is added to the explanation system. Statistical measures such as R2, adjusted R2 and t-tests are used.

5.3 Missing Data points

To answer the third question we have to go back to why we choose the panel data with fixed effects as a quantitative method. We observe the universe of stocks to have some kind of fixed effect varying from company to company but constant in time, which makes it impossible to directly compare the effects of any explanatory variables between them. Since we regard all companies as different unities the ambition is to keep these varying differences between the companies and not to mix them together. There are two easy approaches to fill in missing data, the first one being to fill it in with the arithmetic average over the whole set of companies and the other one to fill in with the arithmetic average over the specific company. The last one is chosen since this one maintains the fixed effect for each company the best.

5.4 Model Testing

The third step in the quantitative analysis is to test all the different twelve variables based models in four tests. All the models have the same target market and do contain all the available data why the following four test environments are created:

1. Swedish small cap (out of sample)
2. Nordic mid/large cap, excluding stocks traded on Stockholm stock exchange (out of sample)
3. Swedish mid/large cap, in sample
4. Swedish mid/large cap, for the 50% and 75% models (out of sample)

To perform the most reliable test possible the data needs to be taken out of sample which means that the model is not tested on data contained in the creation of the model, but since the universe is limited (limit set by available information from Thomson-Reuters) the first two test environments needs to be created. When defining a universe, the first priority is of course to get the largest possible universe, which might intervene with the possibilities to perform out of sample tests. When a model is tested on an environment, which it is not built for, the performance might not be reliable with the possibilities to both show a better and worse performance with the latter having a higher probability. This has to be taken into consideration when evaluating the results. The most reliable test is the fourth one, but this one is of course only applicable to the 50% and 75% models. There are primarily two different qualities that are important for the resulting model, not specifically together, but could show up individually. These are the qualities in predicting a stock to go up or down and to rank the stocks inter relative.

After all the models have been tested, the three models with the best qualities in hit rate of sign prediction are chosen and re-examined. The last, fourth, question above comes into play at this step. When the three best models are chosen, a stepwise regression is brought out based on t tests of each variable. One by one, the variable with the smallest t value is eliminated and the models are tested again. The test results are being closely followed and evaluated on the below mentioned test criterions. The last step in the testing will point out which of these three models perform the best and this one is a final winner and includes the best performing set of explanatory variables.

To evaluate the models the following measures of how well they perform are noted, added to the three mentioned in *5.4.3 The Properties – How Good Is The Model?* ;

1. Sign hit rate – Shows how many times out of the tested projections the real return and the projected return have the same sign.

2. SSR – Sum of Squared residual, which is defined as the difference in predicted return and the actual return.
3. SSR/projection – The above SSR normalized per projection made. Shows how big the errors are in average for each projection.

The focus lies on the sign hit rate since this is the most important for the project over all.

As a last test, the best models, evaluated by the above mentioned criterions will be examined in their ability to rank a group of stocks inter relative. This is measured in the following way; the set of stocks wished to be ranked is divided into three parts with the best 33% on top, the worst 33% in bottom and the rest 33% in the middle. Notice that this is an inter relative ranking and doesn't mean that the bottom part will decrease in value, but only that they will perform not as good as the other 66%. The same test is used for a division into four parts as well. How good the model does the ranking is measured by how many percent of the stocks the model successfully hits the right part with. If a test with a division into three parts is done and the model gets the result 40% then 40% of the ranked stocks are in the correct part comparing projected returns with real returns. A model would add value if it does rank stocks correctly, but misses the sign of the returns. An example is if a model predicts a very exact and fixed return, but to much or too little, then it still can rank the stocks and tell you which ones are forecasted to perform the worst in your portfolio, though it might be hard to point out the edge between sell/buy stocks. Here there is room for more analysis on the side of the model.

6 Results and Discussion

6.1 Step 1 – Simple Regressions

To start with, simple regressions with the full data set were done. This means simple regressions with each factor being the explanatory variable for returns for the whole universe of stocks and with time period 2000-2008. The following transformations showed the highest significance and were used throughout the whole project:

- Price/Book – Negative natural log of Price/Book (P/B)
- Price/Sales – Negative natural log of Price/Sales (P/S)
- Price/Earnings – Negative natural log of Price/Earnings (P/E)
- Dividend yield – Dividend yield powered to five (Div yield)
- Price/cash flow – Negative natural log of Price/Cash flow (P/CF)
- Price – Negative natural log of Price (Low price)
- Return on equity – Original Return on equity (ROE)
- Market capitalization – Negative seventh root of the Market Capitalization (Mark cap)
- Price momentum, 1 month – Positive tenth-log of the current price divided by the prior price (P mom 1)
- Price momentum, 2 months – Positive tenth-log of the current price divided by the prior price (P mom 2)
- Price momentum, 3 months – Positive tenth-log of the current price divided by the prior price (P mom 3)
- Return on equity, 1 year – The Return on Equity one year ago subtracted from the current Return on Equity (ROE mom 1y)

6.1.1 Group 1 & 2

The sector-grouped (group 1) regressions were performed first and since there were only 80 companies in total in the universe, when grouping these into 9 different categories industrials was the largest one with 32 companies followed by industrials and cyclical consumer goods and services sharing second place with 11 companies. The rest of the industries did not cover a large amount of companies and the statistical results varied a lot between them. Actually the only one with a high general statistical significance was industrials with five out of twelve factors having significance above 5 %. Cyclical consumption goods and services had 5 factors significant to a level less than 5 %, but with price momentum for one and two months having a negative sign. Basic Materials, Healthcare, Non Cyclical Consumption Goods and Services and Technologies only had two out of twelve factors having significance above 5 %.

Table 1 regressions with each sector

	P/B	P/S	P/E	Div y.	P/CF	L.P.	P. M. 1	P. M. 2	P. M. 3	ROE. M. 1Y	M.C.	ROE
B.M.	0,02	0,35	0,23	0,53	0,84	2,88**	0,83	-0,44	1,21	1,88*	-0,57	-0,18
Cyclical	2,14*	-0,02	-1,58	1,01	0,65	5,12****	-1,94*^{n}	-3,48***^{n}	29,57****	0,32	4,44****	0,61
Energy	2,09*	1,32	-1,71*^{n}	1,23	1,60	1,90*	0,49	-0,23	-0,85	0,12	1,55	0,14
Financial	2,97**	-1,06	-0,77	1,45	0,16	7,48****	-0,96	-2,60**^{n}	-3,38***^{n}	1,85*	6,18****	1,33
Healthcare	-0,06	0,94	-0,68	1,11	0,04	2,63**	-1,44	0,59	-0,62	-0,63	2,39**	0,25
Industrials	4,92****	-0,55	-0,96	1,62	1,87*	6,37****	-2,17*^{n}	0,64	0,44	6,97****	5,78****	4,98****
Non Cyclical	-0,73	-1,58	-1,37	0,88	1,71*	1,67*	-1,18	-0,02	-0,74	-0,43	1,99*	0,55
Technology	1,32	0,68	0,15	0,73	1,19	3,43***	0,39	-0,66	0,64	0,20	2,32*	0,02
Telecom.	2,02*	0,76	-0,13	1,35	-0,33	2,29*	-2,12*^{n}	-2,29*^{n}	-2,49**^{n}	-0,20	2,37**	-0,92

Significant at the: *5 % level, **1 % level, ***0,1 % level, ****0,01 % level. ^{n} - Negative sign

With only Industrials being the one to some point backed up by statistic results, a model which would be based on which industry the stocks belonged to would not be able to give accurate results if the stock is not Industrial. The lack of data on the majority of the economic sectors reached such low levels that there was no point in continuing with analysing the data in nor smaller time horizons or other time periods. The economic sector sub grouping was rejected. Though these results still contain valuable information. Looking at the averages of the t-values one can see in *Table 2* that low price has the highest statistical significance with a t-value of 3,751 (0,01 % level) followed by market capitalization with a t-value of 2,940 (1 % level) and price momentum 3 months with a t-value of 2,644 (1 % level). Low price is one of the most usual factors indicating a buy and this result clearly proves that it can be used as a signal for Swedish mid/large cap stocks. Market capitalization is of course very correlative with low price, though it gives some effect from the number of shares outstanding in addition to the price. Price momentum for 3 months shows a high predictive power with a positive in-trend effect, in contrast to price momentum for 1 and 2 months, which are though not statistical significant even to a 10 % level but show a negative reversing effect. Together with the last two, price to earnings also show an opposite effect than the hypothesis stated, but also with a very low statistical significance. Low price is also the only one being statistically significant for all sectors, followed by market capitalization (7 sectors), and price/book (5 sectors).

Table 2 average of t-values over all sectors

P/B	1,632
P/S	0,094
P/E	-0,757
Div y.	1,354
P/CF	0,859
L.P.	3,751****
P. M. 1	-0,900
P. M. 2	-0,943
P. M. 3	2,644**
ROE. M. 1Y	1,120
M.C.	2,940**

Significant at the: **1 % level, ****0,01 % level.

The next thing to do was to do the market capitalization grouped regressions. Though, as with the industry grouped regressions, there was a to large problem with a lack of data because of the fact that mid cap stocks suffer from a much less data coverage, due to the fact that they have a shorter lifetime and they are in general less focused on in this sort of databases. As an example of the factor market cap, the mid cap stocks suffer from more than double amount of missing data/stock compared to large cap stocks which could harm the analysis to a great extent since it is important to follow the general rule of statistical significance mentioned above in chapter *4 Data*. With such an unbalance in missing data, models developed for certain groups could differ much in accuracy.

Table 3 simple regressions with mid and large cap

	P/B	P/S	P/E	Div y.	P/CF	L.P.	P. M. 1	P. M. 2	P. M. 3	ROE. M. 1Y	M.C.	ROE
Large Cap	0,98	0,80	0,52	0,89	1,37	8,55****	-1,36	-3,18***^{n}	-1,73*^{n}	1,81*	6,09****	-0,28
Mid Cap	3,12***	1,29	-0,62	1,44	2,02*	8,65****	-1,97*^{n}	-1,84*^{n}	-2,13*^{n}	4,32****	6,72****	3,23***

Significant at the: *5 % level, **1 % level, ***0,1 % level, ****0,01 % level. ^{n} - Negative sign

Three out of twelve of the factors for the large cap are significant to a higher level than 10 % when using a two sided test and eight out of twelve for mid cap. All the momentum factors indicate a reversing trend for 1-3 months in the stock price and they are not correct with the hypothesis. Low price and market capitalization showed a very high statistical significance, as with the sector-grouped regressions, with t-values above 8 and 6 (0,01 % level). Return on equity has a high predicting power for mid cap stocks, both with its fundamental value (t-value 3,23, 0,1 % level) and the momentum based one (t-value 4,32, 0,01 % level), together with price to book (t-value 3,12, 0,1 % level). For large cap the highest in addition to market capitalization and low price is price momentum for 2 months, which is showing a predicting reversing power (t-value -3,18, 0,1 % level). Though, since the significance varied substantially between the different factors, this grouping is, like the industry grouping, rejected.

6.1.2 Group 3-5

When doing the simple regressions with the whole universe and the full time period, all factors except return on equity momentum were significant to a level lower than 5 %.

Table 4 beta value, standard error, t-value and r-square for all factors

	Beta	Std Err	T-value	R2
P/B	0,145	0,019	7,515****	0,021
P/S	0,123	0,019	6,362****	0,015
P/E	0,074	0,019	3,837****	0,006
Div yield	$1{,}71*10^{-8}$	$9{,}861*10^{-4}$	1,732*	0,001
P/CF	0,093	0,019	4,788****	0,009
Low Price	0,187	0,019	9,763****	0,035
P mom 1	0,072	0,019	3,686***	0,005
P mom 2	0,078	0,019	4,032****	0,006
P mom 3	0,062	0,019	3,177***	0,004
ROE mom 1y	0,001	0,001	0,878	0,000
Market cap	0,159	0,019	8,295****	0,025
ROE	$1{,}264*10^{-4}$	$3{,}668*10^{-5}$	3,445***	0,004

Significant at the: *5 % level, ***0,1 % level, ****0,01 % level.

The r-square values in *Table 4* do show a very poor result, but important to note are the t-values. The very strong statistical significance laid a good ground to continue including all the variables in the multiple regressions, including the return on equity momentum. When regressed together, the explanation level might increase. The significance of the variables in a multiple regression is of greater importance than in the simple ones, why this result is enough to let the models 3-5 go to step 2. Again low price and market capitalization were the most statistically significant ones joined by price to book and price to sales.

6.2 Step 2 – Multiple Regressions

When analysing the data set for the years 2006-2008 and 2007-2008 the lack of data was again to large to be able to perform any further analysis with it, with the exception for 3 % data coverage requirement for 2006-2008 and the 50 % data coverage requirement for the same period. Price to cash flow and price to sales were the ones with the least data coverage, both of them with a high statistical significance in the simple regressions. The analysis left was to do multiple regressions with the three sub groups of data, 3-5, for the years 2000-2008, 2003-2008, 2003-2006 and 2006-2008, see *Table 5*.

Table 5 t-values from multiple regressions

	3 % data coverage				75 % data coverage		
	00-08	03-08	03-06	06-08	00-08	03-08	03-06
P/B	1,34	0,00	1,02	2,53**	0,46	0,79	0,80
P/S	1,36	1,34	3,10***	3,45***	-1,18	0,04	-0,80
P/E	1,56	0,81	1,27	0,07	0,33	0,77	0,49
Div Y	1,32	0,51	-0,01	-0,14	-2,38** n	-1,58	-1,75* n
P/CF	-0,08	-0,67	-2,22	0,10	1,49	0,34	-0,36
L.P.	4,76****	0,71	2,11*	1,71*	0,55	-0,17	1,66*
P M 1	-0,09	3,28***	1,75*	2,38**	-0,75	0,36	0,91
P M 2	-1,51	1,12	1,58	0,70	-0,30	1,30	-0,13
P M 3	0,23	0,62	-1,97* n	-2,67** n	0,72	-0,20	-0,70
ROE M 1Y	1,76*	1,38	-0,38	-0,48	1,13	1,31	0,61
M.C.	0,11	2,34**	-0,46	-0,40	0,31	0,37	-0,51
ROE	1,05	0,19	2,13*	-1,04	-0,86	0,82	0,04

	50 % data coverage			
	00-08	03-08	03-06	06-08
P/B	-0,18	-0,71	-0,47	0,54
P/S	0,80	1,62	3,00**	3,44***
P/E	2,36**	2,42**	1,84*	0,77
Div Y	1,39	0,30	-0,01	0,01
P/CF	0,33	-0,90	-1,25	-1,79* n
L.P.	5,25****	2,13*	1,15	3,07**
P M 1	-0,48	2,40**	1,46	0,74
P M 2	1,70*	0,76	1,76*	1,06
P M 3	-1,47	1,35	-1,73* n	-1,23
ROE M 1Y	0,96	0,84	0,44	1,82*
M.C.	-1,71* n	-0,66	-0,30	-1,28
ROE	0,93	0,54	1,00	-1,19

Significant at the: *5 % level, **1 % level, ***0,1 % level, ****0,01 % level. n - Negative sign

The statistical significance from the multiple regressions showed very different results compared to the simple ones, which was expected, since multiple regression deals with how much information is added when including every factor. For example market capitalization is expected to correlate strongly with low price, this is also something which is not allowed due to one of the basic assumptions in the model being all variables are independent of each other, see *Theory 3.4.1 – The model*. With the data set requiring 3 % coverage of data low price seems to have a very strong predicting power when the regression is done with the years 2000-2008, 2003-2006 and 2006-2008, but not at all during the years 2003-2008. This can for example be explained with unbalanced data coverage causing a bias throughout the years. Price momentum for 1 month is, like low price, strong with the years 2003-2008, 2003-2006 and 2006-2008, but very weak 2000-2008. Another interesting result is price to sales being strongly significant with the years 2003-2008 and 2006-2008 and price momentum for 3 months having a strong reversing effect for the same time period, which is a result against our hypothesis. For the data set requiring 75 % data coverage there are fewer significant results with dividend yield being one of the stronger ones but only reaching a t-value of -2,38 for the years 2000-2008, note the negative sign. Low price only shows a t-value of highest 1,66 for the period 2003-2006. For the 50 % data set there are more results showing significance with low price again being the strongest one for the years 2000-2008 with a t-value of 5,25 and 2,13 for the years 2003-2008. Price to earnings is a surprising up comer being significant with a p-value always higher lower than 5 % in all three periods with the highest being t-value of 2,42 for the years 2003-2008. Price momentum for 1 month is also significant with a t-value of 2,40 for the years 2003-2008. A note is the difference between the simple regressions, but low price is still in general a strong predictor in the same time as market capitalization has fallen down and is only significant in the multiple regression in two model periods, with the 50 % data requirement for the years 2000-2008 showing a negative effect on return.

6.3 Performance in Predicting

Enclosing the last part of the project, which one of these models is the best in predicting the right sign of return movements? When comparing the average hit rate over three periods, 2000-2008, 2003-2008 and 2003-2006, the hit rate lied very constant around 50 %, see *Table 6*, which of course is a very negative result, how can a model be used to predict stock movements if one can do the same buy guessing? This result strongly agrees with a market being efficient. Some steps in the project are still left, including the factor elimination and ranking abilities, why the project is brought further. A model which does not predict the right sign can still give information about which stocks are performing the best and which are performing the worst, via ranking.

Table 6 average hit rate

		test env. 1	test env. 2	test env. 3
3 % model				
	00-08	0,511	0,488	0,535
	03-08	0,500	0,471	0,513
	03-06	0,536	0,484	0,532
75 % model				
	00-08	0,524	0,525	0,532
	03-08	0,514	0,516	0,500
	03-06	0,498	0,503	0,513
50 % model				
	00-08	0,535	0,520	0,513
	03-08	0,514	0,416	0,509
	03-06	0,524	0,519	0,531
	06-08	0,531	0,466	0,519

This negative result also pushes hard on that it is necessary with doing an out of sample test within the target market before, which the model is constructed for. According to *Table 6*, the models perform the worst in test environment 2, but also vary a lot within this test. This is because this environment probably is further away in qualities form the target market than 1 and, of course, 3. To choose three models from this result, the average hit rate over test environmnent 1 and 3 is therefore calculated, see *Table 7*.

Table 7 average hit rate over test environments 1 and 3

3 % model		
	00-08	52,277 %
	03-08	50,639 %
	03-06	53,393 %
75 % model		
	00-08	52,805 %
	03-08	50,697 %
	03-06	50,540 %
50 % model		
	00-08	52,438 %
	03-08	51,163 %
	03-06	52,703 %
	06-08	52,510 %

The 3 % model based on the years 2003-2006 is the best performing with an average hit rate of 53,4 %. When the models are this similar in hit rates, the results from a model which can't be tested on an out of sample test are too weak to be supporting the performance of the model, why the 3 % model is rejected. The second best is the 75 % model based on the years 2000-2008 with a hit rate of 52,8 %, third best is the 50 % model based on the years 2003-2006 with a hit rate of 52,7 % and fourth best is the 50 % model based on the years 2006-2008 with a hit rate of 52,5 %.

1. 75% model based on the years 2000-2008
2. 50% model based on the years 2003-2006
3. 50% model based on the years 2006-2008

After stepwise regression on all three with one by one variable eliminated, the variables were taken out in order according to *Table 8*.

Table 8 order of factor elimination for the three best performing models

75 % model, 2000-2008	50 % model, 2006-2008	50 % model, 2003-2006
Price Momentum, 2m	Price Momentum, 1m	Dividend Yield
Market Cap	Return on Equity Momentum	Market Cap
Price/Earnings	Market Cap	Return on Equity Momentum
Price/Book	Price Momentum, 2m	Price/Book
Price Momentum, 3m	Price/Cash flow	Price Momentum,1m
Price Momentum, 1m	Return on Equity	Price/Cash flow
Price/Sales	Price/Sales	Return on Equity
Return On Equity	Price/Earnings	Low Price
Return on Equity Momentum	Dividend Yield	Price Momentum, 3m
Low Price	Price/Book	Price Momentum, 2m
Dividend Yield	Low Price	Price/Sales
Price/Cash flow	Price Momentum, 3m	Price/Earnings

When the twelve variables are put in the same linear system, the ones with the most significant predictability are clearly the low price effect, price/sales, price momentum 3m and return on equity together. Market capitalization is among the ones, which are eliminated quicker, probably because it correlates very highly with price, but the low price effect is stronger than the predictability of market cap, why it is eliminated first. After the stepwise regression was performed, the top three models out of the 36 possible combinations from *Table 8*, when studying hit rates, were the following ones with the listed variables:

Table 9 the best performing models after stepwise regression

75 % model, 2000-2008	75 % model, 2000-2008	50 % model, 2006-2008
• Return on Equity Momentum	• Return On Equity	• Price Momentum, 3m
• Dividend Yield	• Dividend Yield	• Dividend Yield
• Low Price	• Low Price	• Low Price
• Return On Equity	• Price/Cash flow	• Return On Equity
• Price/Sales		• Price/Sales
• Price/Cashflow		• Price/Earnings
		• Price/Book

How well do these ones perform? These models have average hit rates on an out of sample test, test environment 4, according to *Table 10*.

Table 10 average hit rate on test environment 4

	75 % 6 var	75 % 4 var	50 % 7 var
2000-2008	56,1	57,0	53,2

According to *Table 10* the 75 % model with 4 variables gives the highest average hit rate closely followed by the 6 variable version. These two has a higher average hit rate than the 12 variable versions and has improved by stepwise regression, though it is still very low, struggling on mid 50 %. Do they perform different under different market conditions? To find out the period 2003-2004 and 2007-2008 are studied. The first one being the period when the price of the universe of stocks increased the most and the last one when they decreased the most. These are two opposite extreme market conditions and consolidating conditions are put aside.

Table 11 performance of the three best models

(%)		75 % model, 6 var	75 % model, 4 var	50 % model, 7 var
2003-2004	min	60,0	50,0	59,0
	max	80,0	72,0	76,5
	average	65,4	61,0	66,4
2007-2008	min	60,9	63,0	53,8
	max	79,1	76,7	85,7
	average	67,3	69,9	66,0

From *Table 11* one can see that the 75% model with 4 variables is clearly the worst one in the up period with lower average, lower min and lower max. The 75% 6 variable model has a slightly lower average than the 50% model in up but reaches slightly higher both in min and max why this one is taken out as the winner in an up period. In the down period the 75 % model with 6 variables is on second place in all three, min, max and average. The 4 variable model wins in min and average with only a small difference in percent units. The 50% model clearly wins in the max value but is the one, which reaches the lowest in both average and min. The min value in such a model is quite important why the 50% model is rejected. The 4 variable model wins the down period over the 6 variable model, but the difference was not substantially high. If one would use two different models for up and down market and let the user decide with his belief which market to expect and then which model to use he/she would get punished a lot when believing in an down market but ending up with an up market, where the 4 variable model performed much worse than the 6 variable one. This leads us to making a little bit of a sacrifice in hit rate in the down market but instead wins the abandoned possibility for predicting the market up/down movement wrongly. The 75% 6 variable model is the final winner when comparing hit rates.

One last question arises though, the model performs good in an up and down market, but how well does it do in a market which is in a more of a consolidating movement? To show this quality, the model was tested on the period starting in the end of the year 2005 and ending in the summer 2006. The model reaches an average of 56,8%, minimum of 47% and a maximum of 62,5% the model clearly performs worse in small movements in the market. This can be explained with the error the model does, if a stock is moving slightly up; the model could predict it to go up but with the addition of the error term the final predicted sign could still be negative. When there are clear signs of a movement, the model performs well, but worse when the signs are blurrier. One can also look at it through the following perspective, when the model performs the worst and predicts sign only slightly better than one would do with a coin, with an investment in the models recommending one would not loose much, but when following the models recommending in a more volatile market, one could make an investment increase in value. Finally, how well do the model rank stocks? A ranking test was performed on 43 stocks in the beginning in the summer 2007, the beginning of the period, which is referred to as 2007-2008. The model successfully puts 37,2% of the stocks in the right fourth, 51,2% of the stocks in the right third and 76 % in the right half. How well it ranks in an up market can be shown when analysing the beginning of the up period 2003-2004. The model successfully sorts 60% of the 20 stocks in the right half, 35% in the right third and 20% in the right fourth. These tests show that it does have a quality in ranking, at least dividing a set of stocks into the worst and the best performing ones, and the ranking possibility is included in the final program. The statistical results for the 75 % model with 6 variables is included in *Table 12*.

Table 12 statistical results for the 75 % model with 6 variables

Regression statistics	
Multipel-R	0,298
R2	0,089
Adjusted R2	0,073
Standard error	0,037
Observations	363
Sum of diff / prediction (with sign)	4,018%
Sum of absolute diff / prediction	8,308%

	DoF	*Sum of Sq.*
Regression	6	0,0469
Residual	357	0,4805
Totally	363	0,5274

Correlation	*Price/Cashflow*	*Price/Sales*	*ROE*	*Low Price*	*Dividend Yield*	*ROE Mom 1Y*
Price/Cashflow	1					
Price/Sales	0,62	1				
ROE	0,13	-0,24	1			
Low Price	0,62	0,63	-0,42	1		
Dividend Yield	0,13	0,17	-0,10	0,15	1	
ROE Mom 1Y	0,16	-0,05	0,66	-0,16	-0,11	1

	Coefficients	*Standard error*	*t-value*	*p-value*
Constant	0			
Price/cashflow	0,0276	0,0103	2,6692**	0,0080
Price/Sales	-0,0073	0,0065	-1,1156	0,2653
ROE	-0,0006	0,0005	-1,2848	0,1997
Low Price	0,0099	0,0072	1,3790	0,1688
Dividend yield	$-0{,}0031 \cdot 10^{-5}$	0,0000	$-2{,}3823^{**n}$	0,0177
ROE mom 1Y	0,0004	0,0003	1,2587	0,2089

Significant at the: *5 % level, **1 % level. ^{n} - Negative sign

According to the statistical results, see *Table 12*, this model predicts the returns of each stock with an absolute error of 8,308 % and the average error for each prediction is 4,018 %, when average is calculated including sign of all values. The importance of the adjusted R2 can be discussed, since it is a panel data regression, but it reaches 7,3 %. Only two of the 6 variables are statistically significant with a p-value lower than 0,05. They are price to sales and dividend yield. When a multiple regression is performed, these two are said to have the highest predicting power, the other ones not statistically significantly different from zero on a 5 % level. The correlation between the variables reaches the most 0,66, which is between ROE and ROE mom 1y and is not regarded as to high to break the basic assumptions about co-linearity. Finally, the best possible model for predicting stock returns calculates the new price of the stock based on the old price, return on equity, return on equity momentum since 1 year back, price/sales, price/cash flow and dividend yield. The variables affect the price one month ahead in the following fashion:

The lower price today – the higher return
The lower dividend yield today - the higher return
The higher return on equity increase the last year - the higher return
The lower return on equity today - the higher return
The lower price/sales – the higher return
The higher price/cash flow – the higher return

And together they decide the price of a Swedish mid/large cap stock in one month according to the following mathematical relationship, where p_{t+1} is the price of the stock in one month, p_t is the price of the stock today, roemom is the return on equity momentum:

$$p_{t+1} = p_t 10^{0{,}114 + 0{,}028\ln(p/cashflow) - 0{,}007\ln(p/sales) - 0{,}0006(roe) - 0.010\ln(p_t) - 3{,}1 * 10^{-8}(divyield)^5 + 0{,}0004(roemom)}$$

7 Conclusion

After the extensive research and development of a final model, the hypothesis about the efficient market hypothesis is rejected. This project has clearly showed that with only information available for all participants in the market today, one could predict movements of stocks in the future. The final model is not surprisingly based on the highest data requirement of 75 % and performs well in both an up and down market, but worse in a consolidating market. The model reaches as high as 80 % safety in sign hit rate and its usability can be regarded as high, when used in the right way. Classic factors when predicting stock returns, such as low price, market capitalization, price/book value, price/sales, dividend yield and price/cashflow, has in this project proven to contain valuable information for investors. When a model for predicting stock return movements, the statistical properties, such as t-values, are put in second priority after hit rate and a model which can predict ups and downs in returns for the Swedish mid/large cap stock market has been developed. Important is though to see the result as a tool for identifying trading opportunities and not a black-box trading decision maker and apart from the models themselves, the project achieved with laying grounds for a continuation of an even more extensive analysis and future research within the area.

8 Suggestion for Further Research

Focus for further research should lie on increasing the total data set, including both more covered companies (only around 50% of the Swedish mid/large cap stocks were covered) and with a longer time horizon. This is the heaviest improvement to develop more accurate models. A screening of more databases could solve this since the coverage among the many financial databases available differs a lot. If this is done the approach with grouping the universe according to which economic sector the stocks belong to could be continued. A sub grouping according to market capitalization could presumably be done even with only a little bit more data than used in this project. The second point of focus should lie on evaluating and testing the already available data that was used for this project to a higher degree. Even though the models were sorted out, accurate and precise models could have been left out or missed. The third focus point should be to do the project over again with more variables. Many of the documented models do contain substantially more variables than twelve, but how the variables are used is varying from model to model which leads us to a fourth focus point, are there any other smarter ways of using the variables?

9 References

9.1 Books

Jacobs, B. I. & Levy, K. N., (2000), *Equity Management: Quantitative analysis for stock selection,* New York: McGraw Hill.

Haugen, R. A., (2001), *Modern investment theory, 5th ed*, New Jersey: Prentice-Hall.

Bruner, R. P., (1998), *Global Equity Selection Strategy,* Chicago and London: Fitzroy Dearborn Publishers.

Nicholas, J G., (2000), *Market-neutral investing, long/short hedge fund strategies*, Princeton: Bloomberg Press

Bernstein J., (2002), *Momentum Stock Selection, Using the momentum method for maximum profit,* The McGraw Hill Companies

Arrelano M., (2003), *Panel Data Econometrics, Advanced Texts in Econometrics,* Oxford University Press

Dougherty C., (2007), *Introduction To Econometrics*, Third Edition, Oxford University Press

Taleb, N. N., (2007), *The Black Swan*, Great Britain, Allen Lane 2007.

Taleb, N. N., (2004), *Fooled by Randomness*, New York, Thomson/Texere.

9.2 Articles

Johnson G., Ericson S, Srimurthy V., (2007), *The Journal of Alternative Investments*, Institutional Investor Journals.

Armfelt C, Somos D., (2008), *Performance, Benefits and Risks of Active-extension Strategies*, Master Thesis at Stockholm School of Economics.

Cooper M, Gutierrez R. C. Jr, Marcum W., (2001), *On the Predictability of Stock Returns in Real Time*, Journal of Business.

Clarke R, de Silva H, Thorley S., (2001), *Portfolio Constraints and the Fundamental Law of Active Management*, Financial Analysts Journal.

Clarke R, de Silva H, Sapra S and Horley S., (2008), *Long-Short Constraints: How Much Is Enough?,* Financial Analysts Journal.

Sorensen E, Lingjie M, Hua R and Qian E., (2007), *Aspects of Contrained Long/Short Equity Portfolios*, The Journal of Portfolio Management.

Martielli J. D., (2005), *Quantifying the Benefits of Relaxing the Long-Only Constraint*, Worker Paper, SEI Investments Developments.

Ferson W. E. Sarkissian S, Simin T., (1998), *The Alpha Factor Asset Pricing Model: A Parable,* University of Washington.

Jones, P. J., (2008), *Analyzing and Estimating Real Stock Returns*, The Journal Of Portfolio Management, spring edition.

de Bondt G. J., (2008), *Determinants of Stock Prices*, The Journal Of Portfolio Mangement, spring edition.

Alford A, Jones R. C. Lim T, Litterman B., (2004), *Fundamentals Drive Alpha*, Quantitative Equity Research Notes, Goldman Sachs Asset Management.

Burmeister E, Roll R, Ross S. A., Elton E. J., Gruber M. J., Grinold R., Kahn R. N., (march 1994), *A Practitioner's Guide to Factor Models,* The Research Foundation of The Institute of Chartered Financial Analysts.

Bush J. S., (2007), *Value and Growth, Theory and Practice*, Journal of Portfolio Management.

Rapach E. D, Wohar E. M., (2004), *In-Sample vs. Out-of-Sample Tests of Stock Return Predictability in the Context of Data Mining*, Saint Louis University.

Fama, E., (1970), *Efficient Capital Markets: A Review of Theory and Empirical Work*, Journal of Finance.

9.3 Internet

Quantitative Equity Analysis: A Primer on Valuation Models and Backtesting, Team-sys.
http://www.team-sys.com/primer.pdf

Dow, C. G., (1998), *Growth vs. Value Investing*, Dow Publishing Company.
http://www.dows.com/Publications/growth_versus_value_investing.htm

Bäckström U., (2001), *Svensk ekonomi*, Riksbanken.
http://www.riksbank.se/templates/speech.aspx?id=5909

10 Appendix

10.1 Most Discussed Factors

Company specific factors such as:

Value factors

- Book to Price
- Book to Market
- Price to Sales
- Price to Earnings
- Dividend yield
- Earnings Yield of a Trailing Time Period
- Forecasted Earnings Yield
- Trailing Revenue Yield
- Cash flow to Price
- Price
- Share decrease
- Bid-ask spread
- Market capitalization

Growth factors

- Estimate Revision
    - Earnings per share estimate revision
    - Estimated Earnings per Share growth rates
    - Estimate Revision
    - Target Price
- Earnings per Share growth rates
- Earnings surprise
- Relative price strength
- ROE
- Forecast ROE

- Net Margin
- CFPS Growth
- Revenue Growth
- Margin expansion
- Momentum such as:
  - Price Momentum
  - Return on Equity momentum
  - Earnings Momentum
  - Revenue Momentum

Liquidity

- Nbr. of shares traded

Technical

- Moving Average Delta
- Moving Average Convergence/Divergence
- Price volatility’
- Money flow

Macro factors such as:

- Interest rates
- Inflation
- Business activity
- Confidence risk, time horizon risk, inflation risk, business cycle risk, market timing risk
- Change in expected inflation, unexpected inflation, unexpected changes in risk premium, unexpected changes in the term premium
- Economic growth, business cycle, long term and short term interest rate, inflation shock, US Dollar

10.2 Names, RIC-codes and Sectors for All Companies

10.2.1 Swedish Mid/Large-Cap

Name	**RIC**	**RBSS Sector**
Aarhuskarlshamn AB	AAK.ST	Non-Cyclical Consumer Goods & Services
Addtech AB	ADDTb.ST	Industrials
AF AB	AFb.ST	Industrials
Alfa Laval AB	ALFA.ST	Industrials
Atlas Copco AB	ATCOa.ST	Industrials
Axfood AB	AXFO.ST	Non-Cyclical Consumer Goods & Services
Axis AB	AXIS.ST	Cyclical Consumer Goods & Services
Ballingslov International AB	BALL.ST	Industrials
B&B Tools AB	BBTOb.ST	Industrials
Beijer Alma AB	BEIAb.ST	Basic Materials
G&L Beijer AB	BEIJb.ST	Industrials
Betsson AB	BETSb.ST	Technology
Billerud AB	BILL.ST	Basic Materials
Boliden AB	BOL.ST	Basic Materials
Brinova Fastigheter AB	BRINb.ST	Financials
Brostrom AB	BROb.ST	Industrials
Biovitrum AB	BVT.ST	Healthcare
D Carnegie & Co AB	CAR.ST	Financials
Cardo AB	CARD.ST	Industrials
Castellum AB	CAST.ST	Financials
Cloetta Fazer AB	CFAb.ST	Non-Cyclical Consumer Goods & Services
Clas Ohlson AB	CLASb.ST	Cyclical Consumer Goods & Services
Electrolux AB	ELUXb.ST	Cyclical Consumer Goods & Services
Eniro AB	ENRO.ST	Cyclical Consumer Goods & Services
Telefon AB LM Ericsson	ERICb.ST	Technology
Fabege AB	FABG.ST	Financials
Fagerhult AB	FAG.ST	Industrials
Fast Partner	FPAR.ST	Financials
Getinge AB	GETIb.ST	Healthcare
Gunnebo AB	GUNN.ST	Industrials

Heba Fastighets AB	HEBAb.ST	Financials
Hexagon AB	HEXAb.ST	Industrials
Haldex AB	HLDX.ST	Cyclical Consumer Goods & Services
H & M Hennes & Mauritz AB	HMb.ST	Cyclical Consumer Goods & Services
Hoganas	HOGAb.ST	Basic Materials
Holmen AB	HOLMb.ST	Basic Materials
Husqvarna AB	HUSQb.ST	Industrials
Industrial and Financial Systems IFS AB	IFSb.ST	Technology
Intrum Justitia AB	IJ.ST	Industrials
Indutrade AB	INDT.ST	Industrials
Industrivarden AB	INDUa.ST	Financials
Klovern AB	KLOV.ST	Financials
Investment AB Latour	LATOb.ST	Financials
Lindab AB	LIAB.ST	Industrials
Atrium Ljungberg AB	LJGRb.ST	Financials
Lundin Petroleum Ab	LUPE.ST	Energy
Meda AB	MEDAa.ST	Healthcare
Mekonomen AB	MEKO.ST	Cyclical Consumer Goods & Services
Munters AB	MTRS.ST	Industrials
NCC AB	NCCb.ST	Cyclical Consumer Goods & Services
Net Insight AB	NETIb.ST	Telecommunications Services
New Wave Group	NEWAb.ST	Industrials
Nibe Industrier AB	NIBEb.ST	Industrials
Niscayah Group AB	NISCb.ST	Industrials
Nobia AB	NOBI.ST	Industrials
Orc Software AB	ORC.ST	Technology
Q Med AB	QMED.ST	Healthcare
Ratos AB	RATOb.ST	Industrials
Saab AB	SAABb.ST	Industrials
Sandvik AB	SAND.ST	Industrials
SAS AB	SAS.ST	Industrials
Svenska Cellulosa AB	SCAb.ST	Basic Materials
Scania AB	SCVb.ST	Industrials
Securitas Direct AB	SDIRb.ST	Industrials
Seco Tools AB	SECOb.ST	Industrials
Sectra AB	SECTb.ST	Healthcare
Securitas AB	SECUb.ST	Industrials
Skanska AB	SKAb.ST	Industrials

SKF AB	SKFb.ST	Industrials
Skistar AB	SKISb.ST	Cyclical Consumer Goods & Services
SSAB Svenskt Stal AB	SSABa.ST	Basic Materials
Sweco AB	SWECb.ST	Industrials
Swedish Match AB	SWMA.ST	Non-Cyclical Consumer Goods & Services
Tele2 AB	TEL2b.ST	Telecommunications Services
TeliaSonera AB	TLSN.ST	Telecommunications Services
TradeDoubler AB	TRAD.ST	Cyclical Consumer Goods & Services
Trelleborg AB	TRELb.ST	Basic Materials
Tricorona AB	TRIC.ST	Financials
Vostok Gas Ltd	VGASsdb.ST	Energy
Volvo	VOLVb.ST	Cyclical Consumer Goods & Services

10.2.2 Nordic Mid/Large-Cap

Name	**RIC**	**RBSS Sector**
SeaDrill Ltd	SDRL.OL	Energy
Yara International ASA	YAR.OL	Basic Materials
StatoilHydro ASA	STL.OL	Energy
Aker ASA	AKER.OL	Industrials
Norsk Hydro ASA	NHY.OL	Basic Materials
Hafslund ASA	HNA.OL	Utilities
Awilco Offshore ASA	AWO.OL	Energy
StatoilHydro ASA	STO.N	Energy
Orkla ASA	ORK.OL	Non-Cyclical Consumer Goods & Services
DNB NOR ASA	DNBNOR.OL	Financials
Fred Olsen Energy ASA	FOE.OL	Energy
Renewable Energy Corporation ASA	REC.OL	Utilities
Aker Solutions ASA	AKSO.OL	Energy
Sevan Marine ASA	SEVAN.OL	Energy
Marine Harvest ASA	MHG.OL	Non-Cyclical Consumer Goods & Services
Wilhelm Wilhelmsen ASA	WWI.OL	Industrials
Kongsberg Gruppen ASA	KOG.OL	Industrials
Storebrand ASA	STB.OL	Financials
Telenor ASA	TEL.OL	Telecommunications Services
Tandberg ASA	TAA.OL	Cyclical Consumer Goods & Services

Bonheur ASA	BON.OL	Energy
SKAGEN Vekst	SKIVEK.CO	Financials
Petroleum Geo-Services ASA	PGS.OL	Energy
SKAGEN Kon-Tiki	SKIKON.CO	Financials
SKAGEN Global	SKIGLO.CO	Financials
Schibsted ASA	SBST.OL	Cyclical Consumer Goods & Services
Norsk Hydro ASA	NHYDY.PK	Basic Materials
Telenor ASA	TELNY.PK	Telecommunications Services
Frontline Ltd	FRO.OL	Industrials
Outotec Oyj	OTE1V.HE	Basic Materials
Pohjola Pankki Oyj	POH1S.HE	Financials
Stora Enso Oyj	SEOAY.PK	Basic Materials
Kesko Oyj	KESBV.HE	Non-Cyclical Consumer Goods & Services
Sampo Oyj	SAMAS.HE	Financials
Orion Corporation	ORNBV.HE	Healthcare
UPM-Kymmene Oyj	UPM1V.HE	Basic Materials
Metso Oyj	MEO1V.HE	Industrials
YIT Oyj	YTY1V.HE	Industrials
Poyry Oyj	POY1V.HE	Industrials
Cargotec Corporation	CGCBV.HE	Industrials
Outokumpu Oyj	OUT1V.HE	Basic Materials
Rautaruukki Oyj	RTRKS.HE	Basic Materials
Konecranes Oyj	KCR1V.HE	Industrials
Wartsila Oyj Abp	WRT1V.HE	Industrials
Stockmann Oyj	STCBV.HE	Cyclical Consumer Goods & Services
Neste Oil Corporation	NES1V.HE	Energy
Fortum Oyj	FUM1V.HE	Utilities
Nokia Oyj	NOK.N	Technology
KONE Oyj	KNEBV.HE	Industrials
Nokia Oyj	NOK1V.HE	Technology
SanomaWSOY Oyj	SWS1V.HE	Cyclical Consumer Goods & Services
Metso Oyj	MXCYY.PK	N/A
Elisa Oyj	ELI1V.HE	Telecommunications Services
Stora Enso Oyj	STERV.HE	Basic Materials
UPM-Kymmene Oyj	UPMKY.PK	Basic Materials
Nokian Renkaat Oyj	NRE1V.HE	Cyclical Consumer Goods & Services
Sydbank A/S	SYDB.CO	Financials
Carlsberg A/S	CARLb.CO	Non-Cyclical Consumer Goods & Services

Rockwool International AS	ROCKb.CO	Industrials
TDC AS	TDC.CO	Telecommunications Services
De Sammensluttede Vognmaend A/S	DSV.CO	Industrials
Topdanmark A/S	TOP.CO	Financials
H Lundbeck A/S	LUN.CO	Healthcare
AP Moeller Maersk A/S	MAERSKb.CO	Industrials
Danske Bank A/S	DANSKE.CO	Financials
Jyske Bank A/S	JYSK.CO	Financials
Novozymes A/S	NZYMb.CO	Basic Materials
Vestas Wind Systems A/S	VWS.CO	Industrials
Danisco A/S	DCO.CO	Non-Cyclical Consumer Goods & Services
Dampskibsselskabet TORM A/S	TORM.CO	Industrials
FLSmidth & Co A/S	FLS.CO	Basic Materials
Genmab	GEN.CO	Healthcare
Dampskibsselskabet NORDEN A/S	DNORD.CO	Industrials
NKT Holding A/S	NKT.CO	Industrials
Koebenhavns Lufthavne A/S	KBHL.CO	Industrials
Novo Nordisk A/S	NVO.N	Healthcare
BankInvest Hojrentelande	BAIHRL.CO	Financials
TrygVesta A/S	TRYG.CO	Financials
Novo Nordisk A/S	NOVOb.CO	Healthcare
William Demant Holding	WDH.CO	Healthcare
Coloplast A/S	COLOb.CO	Healthcare

10.2.3 Swedish Small Cap

Name	**RIC**	**RBSS Sector**
Sagax AB	SAGA.ST	Financials
Traction AB	TRACb.ST	Financials
Academedia AB	ACADb.ST	Technology
Acando AB	ACANb.ST	Technology
Addnode AB	ANODb.ST	Technology
Aerocrine AB	AEROb.ST	Healthcare
Anoto Group AB	ANOT.ST	Technology
Aros Quality Group AB	AQ.ST	Industrials
Beijer Electronics AB	BELE.ST	Technology

Bilia AB	BILIa.ST	Cyclical Consumer Goods & Services
BioGaia AB	BIOGb.ST	Healthcare
Bioinvent International AB	BINV.ST	Healthcare
Biotage AB	BIOT.ST	Healthcare
Bjorn Borg AB	BORG.ST	Cyclical Consumer Goods & Services
Bringwell AB	BWL.ST	Healthcare
Carl Lamm AB	CLAM.ST	Industrials
Catena AB	CATE.ST	Financials
Cision AB	CISI.ST	Technology
Concordia Maritime AB	CCORb.ST	Industrials
Connecta AB	CNTA.ST	Technology
Corem Property Group AB	CORE.NGM	Financials
Cybercom Group Europe AB	CYBC.ST	Technology
Dagon AB	DAG.ST	Financials
Diamyd Medical AB	DIAMb.ST	Healthcare
Din Bostad Sverige	DIBO.ST	Financials
Dios Fastigheter AB	DIOS.ST	Financials
Elanders AB	ELANb.ST	Industrials
Enea AB	ENEA.ST	Technology
eWork Scandinavia AB	EWRK.ST	Industrials
Fastighets AB Balder	BALDb.ST	Financials
Fenix Outdoor AB	FIXb.ST	Cyclical Consumer Goods & Services
Glycorex Transplantation AB	GTABb.NGM	Healthcare
Hemtex AB	HEMX.ST	Cyclical Consumer Goods & Services
HiQ International AB	HIQ.ST	Technology
HL Display AB	HLb.ST	Industrials
HMS Networks AB	HMSN.ST	Technology
IBS AB	IBSb.ST	Technology
Impact Coatings	IMPC.ST	Industrials
ITAB Shop Concept AB	ITABb.ST	Industrials
KABE AB	KABEb.ST	Cyclical Consumer Goods & Services
Karo Bio AB	KARO.ST	Healthcare
Know IT AB	KNOW.ST	Technology
Lagercrantz Group AB	LAGRb.ST	Industrials
Lappland Goldminers AB	GOLD.ST	Basic Materials
LBI International AB	LBI.ST	Technology
Malka Oil AB	MALK.ST	Energy
Medivir AB	MVIRb.ST	Healthcare

Micronic Laser Systems AB (Publ)	MICR.ST	Industrials
Midway Holding AB	MIDWb.ST	Cyclical Consumer Goods & Services
Morphic Technologies AB	MORPb.ST	Industrials
Nederman Holding AB	NMAN.ST	Technology
NeoNet AB	NEO.ST	Financials
Net Entertainment NE AB	NETb.NGM	Technology
Nolato AB	NOLAb.ST	Basic Materials
NOTE AB	NOTE.ST	Technology
Oasmia Pharmaceutical AB	OASMa.NGM	Healthcare
OEM International AB	OEMb.ST	Industrials
Opcon AB	OPCO.ST	Cyclical Consumer Goods & Services
Orexo AB	ORX.ST	Healthcare
Poolia AB	POOLb.ST	Industrials
Pricer AB	PRICb.ST	Industrials
Proffice AB	PROEb.ST	Industrials
Raysearch Laboratories AB	RAYb.ST	Healthcare
Rederi AB Transatlantic	RABTb.ST	Industrials
Rejlers	REJLb.ST	Technology
rnb Retail and Brands AB	RNBS.ST	Cyclical Consumer Goods & Services
Russian Real Estate Investment Company AB	RURIb.ST	Financials
Semcon AB	SEMC.ST	Industrials
Sensys Traffic AB	SENS.ST	Industrials
Sintercast AB	SINT.ST	Industrials
Skanditek Industriforvaltning AB	STEK.ST	Financials
Studsvik AB	SVIK.ST	Industrials
Svedbergs i Dalstorp AB	SVEDb.ST	Cyclical Consumer Goods & Services
Svolder AB	SVOLb.ST	Financials
Swedol AB	SWOLb.ST	Industrials
VBG Group AB	VBGb.ST	Industrials
VLT AB	VLTb.ST	Cyclical Consumer Goods & Services
XANO Industri AB	XANOb.ST	Industrials
Zodiak Television AB	ZODIb.ST	Cyclical Consumer Goods & Services

10.3 Market Capitalization and Sector Status

10.3.1 Market Capitalization

Date	Total	S-Cap	M-Cap	L-Cap
2000-04-03	26	2	12	12
2000-07-03	33	6	15	12
2000-10-03	32	4	15	13
2001-01-03	36	9	14	13
2001-04-03	58	20	21	17
2001-07-03	61	20	23	18
2001-10-03	62	21	23	18
2002-01-03	64	21	23	20
2002-04-03	62	19	23	20
2002-07-03	69	25	23	21
2002-10-03	69	26	27	16
2003-01-03	70	24	28	18
2003-04-03	69	24	27	18
2003-07-03	70	24	26	20
2003-10-02	70	21	28	21
2004-01-05	71	21	27	23
2004-04-05	72	19	30	23
2004-07-05	72	18	31	23
2004-10-05	72	16	32	24
2005-01-05	72	15	30	27
2005-04-05	72	12	32	28
2005-07-05	72	12	30	30
2005-10-05	72	9	31	32
2006-01-05	75	6	36	33
2006-04-05	75	5	33	37
2006-07-05	76	6	33	37
2006-10-05	79	5	34	40
2007-01-05	80	3	34	43
2007-04-05	80	1	38	41
2007-07-05	80	1	38	41
2007-10-04	80	1	40	39
2008-01-02	80	0	41	39
2008-04-02	80	1	43	36

10.3.2 RBSS Sectors

	Nbr of Companies
Basic Materials (B.M.)	8
Cyclical	11
Energy	2
Financial	11
Healthcare	5
Industrials	32
Non Cyclical	4
Technology	4
Telecommunication	3