Factor Analysis of a Low Market Beta Portfolio in the Nordics

ARVID ORBACK

MAGNUS NORDLINDER
Factor Analysis of a Low Market Beta Portfolio in the Nordics

ARVID ORBACK

MAGNUS NORDLINDE
Abstract

The return of publicly traded assets has been studied by both academia and commercial institutions, using models with different sets of factors. Building on the work of previous results in this field, such as the CAPM-model, the three-factor model by Fama and French, and the four-factor model by Carhart, this thesis studies the return of a low market beta portfolio in the Nordic stock market. This is done using multiple linear regression on different risk factors that take into account volatility, company size, book-to-market ratio, and momentum. The choice of factors represents different risks in the market. Results of the thesis find that half of the variation of returns is explained by the chosen model.
Sammanfattning

Acknowledgements

We would like to thank Nordnet Bank for providing guidance, literature articles and data. Especially, we would like to thank Gustaf Haag, at Nordnet, who's supervisory has been of great help.
Author

Arvid Orback, aorback@kth.se
Magnus Nordlinder, mnordli@kth.se
Industrial Engineering and Management
KTH Royal Institute of Technology

Place for Project

Stockholm, Sweden
KTH and Nordnet Bank

Examiner

Professor of Mathematical Statistics, Henrik Hult,
Department of Mathematics
KTH Royal Institute of Technology

Supervisors

Professor of Mathematical Statistics, Boualem Djehiche
Department of Mathematics
KTH Royal Institute of Technology

Professor of Economics, Hans Lööf
Department of Industrial Economics and Management
KTH Royal Institute of Technology

Portfolio Manager, Gustaf Haag
Nordnet AB
## Contents

1 **Introduction** 1  
1.1 Background ............................................. 1  
1.2 Purpose and Problem Formulation ......................... 1  
1.3 Research Question ....................................... 2  
1.4 Scope and Limitations ................................... 2  
1.5 Feasibility ............................................. 3  

2 **Economic Theory** 3  
2.1 CAPM .................................................. 3  
2.2 Three-Factor Model ..................................... 4  
2.3 Four-Factor Model ...................................... 4  
2.4 Five-Factor Model ...................................... 5  
2.5 Factors ................................................ 5  
2.5.1 Alternative Risk Premia (ARP) ...................... 5  

3 **Mathematical Theory** 6  
3.1 Residual Analysis ...................................... 6  
3.1.1 Residual Scaling ..................................... 7  
3.1.2 Residual Plots ....................................... 8  
3.2 Diagnostics for Leverage and Influence ................ 10  
3.2.1 Hat Matrix ......................................... 10  
3.2.2 DFFITS ............................................ 10  
3.2.3 Cook’s Distance .................................... 11  
3.2.4 COVRATIO .......................................... 11  
3.3 Multicollinearity ...................................... 12  
3.3.1 Plot of Regressor vs Regressor ...................... 12  
3.3.2 Correlation Matrix .................................. 13  
3.3.3 Variance Inflation Factors ......................... 13  
3.4 Quality Measures ..................................... 13  
3.4.1 Model Evaluation Criteria ......................... 13  
3.4.2 Hypothesis Testing .................................. 15  
3.4.2.1 Test for Significance of Regression ............ 15  
3.4.2.2 Tests on Individual Regression Coefficients .... 15  
3.4.3 Confidence Interval on Estimated Coefficients in Multiple Regression ..................... 16  
3.4.4 Cross-Validation .................................. 16  

4 **Data** 17  
4.1 Description of Data ..................................... 17  
4.2 Data Collection ......................................... 17
5 Methodology
   5.1 Literature Study ................................................. 17
   5.2 Construction of Factors ........................................ 18
      5.2.1 Description of Factors .................................... 18
      5.2.2 Regressor Variables ....................................... 18
      5.2.3 Response Variable ......................................... 20
   5.3 Model for Testing ................................................ 20

6 Results
   6.1 Financial Returns ............................................... 20
   6.2 Multiple Linear Regression Analysis ............................. 21
      6.2.1 Residual Analysis .......................................... 21
      6.2.2 Influence and Leverage ..................................... 24
      6.2.3 Adjustments for Influence ................................ 29
      6.2.4 Multicollinearity .......................................... 29
      6.2.5 Model Characteristics ..................................... 30

7 Discussion
   7.1 Limitations and Method Critique ................................. 35
   7.2 Mathematical Approach ......................................... 35
   7.3 Aspects of Industrial Economics ................................. 38
      7.3.1 Factor Classification ..................................... 39
   7.4 Further Research ................................................ 40

8 Conclusion .......................................................... 40

9 Appendix A .......................................................... 42

10 Appendix B ........................................................ 45
1 Introduction

1.1 Background
To understand and explain the pricing of assets in conditions of risk, financial researchers developed the first model to explain the expected return of an asset in 1964. The capital asset pricing model (CAPM), introduced independently by John Lintner, Jack Treynor, and William F. Sharpe, is a linear model and dependent only on the market return and the risk-free rate.*[17]

Since then, several new models have been composed, extended with new factors. Today, the factor approach has become an effective mean for portfolio managers in finding different systematic approaches to diversify a portfolio. One or several factors, in a specific region or specific sector, can be chosen to select a portfolio with high correlation to specific risk premia.

Based on previous research, Carhart (1997) composed a model with four factors describing market return, size of a firm, book value† and momentum [3]. This thesis examines how the return of a low market beta‡ portfolio can be described by Carhart’s model, with an additional volatility factor.

In specific, the volatility factor and the momentum factor (the latter a part of Carhart’s model) are often categorized as alternative risk premia, denoted ARP. These premia hold certain characteristics that will be further discussed in the thesis. Since a low market beta portfolio is not explained by the market return, it is of interest to analyze the impact of the remaining risk factors used in the model.

1.2 Purpose and Problem Formulation
As discussed in the background, the purpose of the thesis is to perform a regression analysis on a portfolio, to find out if the return can be described by alternative risk premia. Throughout the thesis, the terms ”equity risk factors” and ”alternative risk premia” are used interchangeably.

---

*The risk-free interest rate, denoted $R_f$, is the rate of return of a hypothetical investment without risk of financial loss. This rate is assumed to be the inter-bank rate.[14, p. 98]
†The book value is the net asset value of a company. It is equal to a firm’s total assets minus its intangible assets and liabilities.[14, p. 59]
‡Beta ($\beta$) considers the correlation of a stock against some type of benchmark, typically the most common index of the country where the stock is listed.

$$\beta_i = \frac{Cov(R_i, R_p)}{Var(R_p)}$$

where $R_i$ is the return of asset $i$ and $R_p$ is the return of the reference portfolio. More precisely, beta is the expected percentage change in its return given a 1% change in the return of the underlying market portfolio. In CAPM, $R_p$ is referred to as $R_m$, the return of the market.[14, p. 375]
The studied portfolio is composed so that it has a low correlation with the market. This thesis will in turn analyze how the low-beta* portfolio correlates to other risk equity factors. In other words, the portfolio strives towards zero exposure to the market and instead exposure to alternative risk premia.

The thesis will in particular study the correlation with two alternative risk premia, momentum and volatility. To accomplish this, the thesis expands Carharts four factor model with adding a volatility factor.

The thesis in question and the subject on which it relies is relevant for several reasons. The research of alternative risk premia is ongoing both in academic and non-academic research, but the subject has not yet been widely established [15, p.3]. Furthermore, the research is closely linked to the achievements behind CAPM and the subsequent multi-factor models. It is possible that alternative risk premia and further research in equity risk factors could be a part of the expansion of the stock return models in the future.

1.3 Research Question

Based on the background and purpose, the following research question forms the basis of the thesis:

*How does a set of risk factors correlate with the return of a low market beta portfolio?*

1.4 Scope and Limitations

The universe of the portfolio is set to stocks on the Swedish, Norwegian, Finnish and Danish stock markets. The limitation in geography is motivated by geographical influence and data sufficiency.

Since culture has an effect on the human bias in stock trading, a universe of countries with resembling markets should be chosen. The economy of the countries should not respond too differently on macro effects. Countries too distant from each other may also be affected by surrounding events differently. Furthermore, by including stocks from more than one country, a larger data set of stocks can be analyzed.[15, p.52]

The portfolios will be constructed and analyzed monthly during a five-year period (2014-2019). The length of period was determined based on the suggestion of Bartholdy and Peare (2005) who stated that the estimate for beta would include more bias with longer time periods [2]. Furthermore, macro events and

---

*Henceforth, when low-beta or low market beta is mentioned, this refers to the correlation of the portfolio to the market return, i.e. traditional market beta (β) as described by Sharpe [17].
changes in economic cycles often have more influence during longer time periods. The effects of these constitute great impact on stocks and are desired to be reduced for the thesis.

The stated universe is limited by a market capitalization criterion. This criterion excludes stocks not traded on the small, mid or large cap of each country’s stock exchange. Stocks with low market capitalization often infer low trading activity. If these stocks were to be included, results would be more biased since these stock prizes often increase or decrease unjustly due to few transactions. As such, the thesis excludes stocks traded on equity markets such as First North.[2]

1.5 Feasibility

Previous and ongoing research about factor analysis has been favorable for the thesis in question; defined and easily adaptable methods have already been created. It is however more difficult to find information about how ARP factors should be defined, since less academic research has been made on this topic.

A large set of data of monthly returns of over 800 stocks over a 5-year period (i.e. 60 months) has been collected. The high number of separate returns have in turn composed 60 accurate observation points. As many methods of regression analysis perform sufficiently enough with this number of data points, the results are considered credible.

2 Economic Theory

In order to fully encompass the aim of the thesis, it is recommended to be acquainted with the development of the different economic models. Below, the models are presented in subsequent order, starting from the capital asset pricing model (CAMP). This is followed by a thorough description of alternative risk premia and a brief statement of relevant factors.

2.1 CAPM

The capital asset pricing model was introduced independently by John Lintner, Jack Treynor, and William F. Sharpe (1964). CAPM uses only one factor to explain the expected market return of an asset. The model is stated as follows:

\[
E[R_i] = R_f + \beta_{m,i} (E[R_m] - R_f) + \alpha
\]

where \(E[R_i]\) is expected return of asset \(i\), \(R_m\) is the return for the whole market of assets, \(R_f\) is the risk-free rate, and \(\beta_{m,i}\) is the coefficient representing the market sensitivity. The regression coefficient \((\beta_{m,i})\) measures the asset’s movement in relation to the market. CAPM is used by corporations in measuring market capital cost and as a basis for making financial decisions.[17]
2.2 Three-Factor Model

Fama and French (1992) improved the CAPM model by introducing the variables ME (market equity) as a measure for size, and BE/ME (book equity to market equity) as a measure for value - both, factors which were not captured by the single market factor. Assuming that corporations with low market capitalization have higher returns than large market capitalization corporations, and that corporations with low book equity to market equity are valued less than corporations with high BE/ME ratio, Fama and French showed that the three-factor model was superior to the CAPM model for different portfolios of assets.[7]

\[
E[R_i] = R_f + \beta_{m,i} (E[R_m] - R_f) + \beta_{s,i} SMB + \beta_{v,i} HML + \alpha
\]  

(2)

In addition to the variables in the CAPM model, Fama and French poses the two additional variables:

- SMB (Small Minus Big)
- HML (High Minus Low)

with corresponding coefficients \( \beta_{s,i} \) and \( \beta_{v,i} \), \( s \) and \( v \) for size respectively value.

2.3 Four-Factor Model

Carhart (1997) extended the three-factor model by adding a fourth factor describing the momentum of the underlying asset. The momentum was measured using a portfolio long on stocks that performed the best during a period, and a portfolio short on stocks that performed the worst during the same period.

Carhart’s model was tested on mutual funds from 1963 to 1993, constructed in to ten different equal-weighted portfolios based on previous calendar year returns. The model was able to explain a larger extent of the variation in excess returns than both the CAPM and the three-factor model could.[3]

The model is stated as follows:

\[
E[R_i] = R_f + \beta_{m,i} (E[R_m] - R_f) + \beta_{s,i} SMB + \beta_{v,i} HML + \beta_{m,i} UMD + \alpha
\]  

(3)

In addition to the variables of the three-factor model by Fama and French, a factor measuring momentum is introduced in Carhart’s model:

- UMD (Up Minus Down)

with the corresponding coefficient \( \beta_{m,i} \), \( m \) for momentum. MOM is used analogously with UMD.
2.4 Five-Factor Model

Carhart’s four-factor model was challenged by Fama and French (2014) whom instead expanded their previous three-factor model to a five factor model, with two factors describing profitability and investments [5]. Fama and French still did not use momentum as a factor. The five-factor model is stated as follows:

\[ E[R_i] = R_f + \beta_{m,i} (E[R_m] - R_f) + \beta_{s,i}SMB + \beta_{v,i}HML + \beta_{p,i}RMW + \beta_{inv,i}CMA + \alpha \]  

(4)

In addition to the variables of the three-factor model, Fama and French pose the two additional variables:

- RMW (Robust Minus Weak)
- CMA (Conservative Minus Aggressive)

with corresponding coefficients \( \beta_{p,i} \) and \( \beta_{inv,i} \), \( p \) and \( inv \) for profitability respectively investments.

Since the last model was published, several additional studies and research have been made, both by academic and financial institutions. In addition to validating the different models, more factors have been added in attempts to better explain the expected return.[15]

2.5 Factors

Factors are constructed based on historical observations. They are composed in such a way that they convey the difference of performance between characteristics. Fama and French, for example, stated that small size firms tend to outperform big size firms over time. In addition, Fama and French meant that growth stocks (i.e. high book-to-market\(^*\) ratio) achieve higher returns than value stocks (i.e. low book-to-market ratio) over time.[8]

2.5.1 Alternative Risk Premia (ARP)

Expansion of factor models are often done by attempting to capture smart beta (i.e. expanding the original market beta by reducing alpha\(^†\)). A later field of interest has been in so called alternative risk premia, abbreviated ARP.[16, pp. 7-8]

\(^*\)The book-to-market ratio is the ratio between a firm’s market value and book value. A low value indicates a value stock, i.e. a case of overvaluation. A high value indicates a growth stock, i.e. a case of undervaluation.[14, pp. 61-62]

\(^†\)Alpha (\( \alpha \)) is a term that describes the ability of a stock or portfolio to beat its underlying benchmark e.g. the market. A positive alpha indicates outperformance, a negative alpha describes the opposite.[14, p. 448]
A traditional risk premia is the equity market exposure. Simply stated, the risk that comes with this market exposure potentially rewards the investor. Alternative risk premia can simply be described as factors that are not considered traditional. Since it is a relatively new area, definitions of alternative risk premia sometimes differ, but widely discussed is that the premia hold some special and valuable qualities that might be interesting in explaining the excess return of an asset. Main attributes of alternative risk premia are that they aim to reduce the underlying asset’s exposure to the market, while at the same time capture a premia. These premia can, for example, be based on low-valued stocks, small firms or low volatile stocks. Unlike traditional risk premia, such as the market factor, ARP have the ability to capture different types of human bias.[15, p. 3]

The characteristics of ARP act as a tool that creates a different type of risk. This risk induces possible capital gains while at the same time diversifies the portfolios of investors. ARP frequently uses long and short positions* to reduce exposure to certain factors while increasing exposure to the specific ARP.[15, p. 3]

Two commonly stated alternative risk premia are momentum and volatility. Research about factor models state that momentum contribute with value in explaining returns and bases on long term stock performance [6][3]. Volatility is also a factor that can capture underlying risk and premia. For example, an investor might choose to invest in low-volatile stocks to decrease risk while at the same time capture a premia [1].

3 Mathematical Theory

For increased understanding of the results, used methods and criteria are described in the section of mathematical theory. Some basic knowledge about multiple linear regression and underlying conditions that must be met are presented in Appendix B. Before reading the subsequent part, this theory should be understood.

3.1 Residual Analysis

Residual analysis is a procedure to detect unmotivated deviating points in the y space (called outliers) and other model inadequacies.

Residual Definition

Residuals are designed as follows:

$$e_i = y_i - \hat{y}_i, \quad i = 1, ..., n$$

*The term long means to buy an asset, expecting it to increase in value over time. The term short means to sell an asset that the investor do not own (interpret as borrowed asset), with the intention to buy it at a future price, hopefully lower than today’s price.[14, p. 405]
That is, a residual is the distance from the observed value to the corresponding value of the fitted line. It may be interpreted as the variability of the response that is not captured by the fitted model, or as the visible errors manifested. Residuals have zero mean and a variance of $\sigma^2$. The approximate average variance can be calculated by

$$\frac{\sum_{i=1}^{n}(e_i - \bar{e})^2}{n - p} = \frac{\sum_{i=1}^{n}e_i^2}{n - p} = \frac{SS_{Res}}{n - p} = MS_{Res}$$

(6)

where $n - p$ is the degrees of freedom.[4, p. 130]

### 3.1.1 Residual Scaling

Residual scaling can be extra helpful in finding outliers and extreme values, i.e. data points that are unjustly distant in the y-space. This is because it holds the property of exposing hidden extreme values.

#### Standardized Residuals

The standardized residuals hold a zero mean and an approximate unit variance. A large standardized residual, $d_i > 3$, potentially indicates an outlier.[4, pp. 130-131]

$$d_i = \frac{e_i}{\sqrt{MS_{Res}}}, \ i = 1, 2, ..., n$$

(7)

#### Studentized Residuals

By standardizing the ith residual by the exact standard deviation, the residual scaling is improved. The exact standard deviation may be derived from

$$e = (I - H)y$$

(8)

where

$$Hy = \bar{y}$$

(9)

Given that the hat matrix is symmetric ($H' = H$) and idempotent ($HH = H$), $I - H$ also is


$$= I - H - H + H =$$

$$= I - H$$

(10)

By substituting $y = X\beta + \epsilon$ in equation 10, the following is yielded

$$e = (I - H)(X\beta + \epsilon) = X\beta - HX\beta + (I + H)\epsilon =$$

$$= X\beta - X(X'X)^{-1}X'X\beta + (H - I)\epsilon =$$

$$= (H - I)\epsilon$$

(11)

The covariance matrix thus becomes

$$Var[e] = Var[(I - H)e] = (I - H)Var[e](I - H)' = \sigma^2(I - H)$$

(12)
since \( \text{Var}[\epsilon] = \sigma^2 I \). The variance of the \( i \)th residual becomes

\[
\text{Var}[\epsilon_i] = \sigma^2 (1 - h_{ii})
\]  

(13)

where \( h_{ii} \) is the diagonal elements of the hat matrix and \( 0 \leq h_{ii} \leq 1 \). The studentized residuals thus is written as

\[
\begin{aligned}
 r_i &= \frac{e_i}{\sqrt{MS_{\text{Res}} (1 - h_{ii})}}, \quad i = 1, 2, \ldots, n \\
\end{aligned}
\]

(14)

Since \( h_{ii} \) describes a measure of location in \( x \) space, the variance of residuals is dependent on the location of the observation point in question. This implies that observations further away from the centroid have smaller variances (\( h_{ii} \) increases) than those near the center (\( h_{ii} \) decreases).

As distant points tend to pull the fitted line against them and simultaneously possess low variances, these observations are generally hard to detect. By dividing the residual by the exact correlating standard deviation, all variances of the residuals becomes 1 and hidden extrapolated points* may easier be detected.[4, pp. 131-133]

**Externally Studentized Residuals (R-Student)**

R-student residuals estimate \( \sigma^2 \) by relying on a data set with the \( i \)th observation removed. This yields

\[
\begin{aligned}
 t_i &= \frac{e_i}{\sqrt{S_{(i)}^2 (1 - h_{ii})}}, \quad i = 1, 2, \ldots, n \\
\end{aligned}
\]

(15)

where

\[
S_{(i)}^2 = \frac{(n - p)MS_{\text{Res}} - e_i^2/(1 - h_{ii})}{n - p - 1}, \quad i = 1, 2, \ldots, n
\]

(16)

Here, \( S_{(i)}^2 \) is the estimate of \( \sigma^2 \) with the \( i \)th observation removed.[4, p. 135]

**3.1.2 Residual Plots**

Graphical analysis of residual plots are very effective in finding deviating observations and to verify the underlying assumptions that have to be met.[4, p. 136]

**Normal Probability Plot**

*Extrapolation is the problem of estimating beyond the original observation region, called the regressor variable hull (RVH). If the set of data points constitutes an ellipsoid in space, an alarming data point might be disregarded as it exists outside the geometry of the points but inside its permitted range.

Points distant might severely influence the fitting of the regression model, misleadingly making the observer believe that the point holds a very low variance (whereas the variance of points near the centroid is large).[4, pp. 107,110]
Large deviation from the normality assumption will have a negative effect on the regression analysis. If this occurs, big errors will appear in the t and F statistics, as well as in confidence intervals, since these depend on the normality assumption. In addition to this, non-normality might result in heavier tails than normal, which makes the least-squares very sensitive to small sub-sets of data. Outliers produced by the wrong distributions may in fact ”pull” the least-squares fit too much in a certain direction.

A method to confirm whether the normality assumption holds or not, is to plot a normal probability plot of the residuals. Given that the normality assumption holds, the graph will present a straight line. In order for the results to be accurate, a number of about 20 observations is required.[4, pp. 138-139]

**Plot of Residuals vs Fitted Values**

By plotting the externally studentized residuals versus the fitted values, it is possible to draw conclusions about how the variance of the errors behave. If patterns show anything but a horizontal band, the variance of the error is not constant (as it should be). It might be that the variance is an increasing or decreasing function of y, that the model errors follow another distribution than the normal or simply that the relationship between the response and some regressors are non-linear.

In excess of being able to constitute the need of a transformation, the plot may reveal abnormally large residuals, i.e. potential outliers.[4, pp. 139-140]

**Plot of Residuals vs Regressor**

With the same approach as for residuals against the fitted values, but now exchanging the fitted values with regressor j (i.e. horizontally scaled with \( x_{ij} \)), it is possible to check the pattern of the error variances. If this pattern is other than horizontal, the assumed relationship between y and the regressor \( x_i \) is not correct, e.g. it might be of a higher order than linear.[4, p. 141]

**Partial Regression Plots**

Partial regression plot, also called adjusted-variable plot, resembles the plot of residuals versus fitted values. The main difference is that this plot considers how a single regressor is influenced by the other regressors, i.e it studies the marginal effect of a regressor given the remaining regressors in the model. Here, \( y \) and \( x_j \) are both regressed against the other regressors in the model and the residuals obtained from each regression procedure.

Apart from showing pattern of variance, the plot will foremost provide information about a regressor’s marginal usefulness in the regression, i.e. state its individual contribution to the model. This is important while interdependencies between regressors might cover individual characteristics.[4, p. 143-144]
3.2 Diagnostics for Leverage and Influence

A leverage point is a point remote in the x space. This point is not necessarily misleading the regression analysis, it might as well be consistent with the rest of the observations. If this is the case, estimated coefficients will not be influenced substantially. However, it will most likely have an effect on model statistics such as $R^2$ and MSE.

An influential point may not only be remote in terms of the specific values of the regressors (outliers), the observed response may also deviate in the x space (leverage). Influence point can be described as points that are not consistent with the values that would have been predicted based on the rest of the observations.[4, p. 211].

3.2.1 Hat Matrix

The hat matrix

$$H = X(X'X)^{-1}X'$$

(17)

describes the impact each response value has on the fitted line. It is derived from the least squares estimator as

$$\hat{y} = X\hat{\beta} = X(X'X)^{-1}X'y = Hy$$

(18)

The diagonal elements in the hat matrix, $h_{ii} = x_i(X'X)^{-1}x_i'$ (where $x_i$ is the $i$th row of $X$), are useful in detecting leverage points. The elements describe the Euclidean distance of the regressors from the centroid as well as the density of the point in the regressor variable hull, i.e. the smallest convex set encompassing all data points.

An observation is often considered a leverage point if it has a hat value $h_{ii}$ larger than $2p/n$. It is recommended that this analysis is done in conjunction with a method that detect potentially outliers.[4, pp. 212-213]

3.2.2 DFFITS

Since it is often necessary to both look at leverage and outliers in order to detect influential point, some methods are constructed to encompass both aspects.

$DFFITS_i$ is a deletion diagnostic. This means that it is based on the removal of a certain point. The formula is presented as below

$$DFFITS_i = \frac{\hat{y}_i - \hat{y}_{(i)}}{\sqrt{s^2_{(i)}h_{ii}}}, \ i = 1, 2, ..., n$$

(19)

where $h_{ii}$ is the diagonal element of the hat matrix and $\hat{y}_{(i)}$ is the fitted value at index i if the $i$th observation was removed.
$DFFITS_i$ can be interpreted as the number of standard deviations that the fitted value $\hat{y}_i$ changes if observation $i$ is removed. To better understand the diagnostic, it may be rewritten as

$$DFFITS_i = t_i \left( \frac{h_{ii}}{1 - h_{ii}} \right)^{1/2}$$  \hspace{1cm} (20)$$

where $t_i$ is the R-student residual. By looking at equation 9, one can see that $DFFITS_i$ consists of one part detecting leverage, i.e. $[h_{ii}/(1 - h_{ii})]^{1/2}$, and one part detecting outliers, i.e. the R-student residual $t_i$.

Belsley, Kuh, and Welsch (1980) suggests that values that answer $|DFFITS_i| > 2 \sqrt{p/n}$, where $p$ is the number of regressors and $n$ is the number of observations, warrants attention (i.e. they might be influential points).[4, pp. 217-218]

3.2.3 Cook’s Distance

Another deletion diagnostic that considers both the location of the point in the $x$ space and the $y$ space, is Cook’s distance. This measure is based on the squared distance between the least-squares estimate $\hat{\beta}$ based on all observations and the least-squares estimate $\hat{\beta}_{(i)}$ based on all points except the $i$th one. Mathematically, Cook’s distance is presented as

$$D_i = D(X'iX', pMS_{Res}) = \frac{(\hat{\beta}_{(i)} - \hat{\beta})'X'X(\hat{\beta}_{(i)} - \hat{\beta})}{pMS_{Res}}, \text{ } i = 1, 2, ..., n$$ \hspace{1cm} (21)$$

where $p$ is the number of regressors and $MS_{Res}$ is the approximated residual sum of squares.

The ability to both measure the location of an observation in $x$ space and its residual magnitude may be easier to interpret when the measurement is rewritten on the form

$$D_i = \frac{r_i^2}{pVar[\hat{y}_i]} = \frac{r_i^2}{p} \frac{h_{ii}}{1 - h_{ii}}, \text{ } i = 1, 2, ..., n$$ \hspace{1cm} (22)$$

Points with large values of $D_i$ have considerable influence on the least-squares estimates $\hat{\beta}$. A value above 1 warrants attention.[4, pp. 215-216]

3.2.4 COVRATIO

The measure of $COVRATIO_i$ provides information about the overall precision of estimation. The effect of the $i$th observation of the general precision may be written as

$$COVRATIO_i = \frac{|(X'_{(i)}X_{(i)})^{-1}s^2_{(i)}|}{|X'X|^{-1}MS_{Res}}, \text{ } i = 1, 2, ..., n$$ \hspace{1cm} (23)$$
or rewritten as

\[ \text{COV RATIO}_i = \frac{(S^2_{(i)})^p}{MS_{Res}^p} \left( \frac{1}{1 - h_{ii}} \right), \quad i = 1, 2, ..., n \]  

(24)

If \( \text{COV RATIO}_i > 1 \), the \( i \)th observation contribute to the precision of estimation (values less than unity degrades the precision). More precisely, Belsley, Kuh, and Welsch (1980) stated that values that satisfies \( |\text{COV RATIO}_i| > 1 - 3p/n \) (where \( p \) is the number of used regressors and \( n \) is the number of observations) are influential. More precisely, large values of \( \text{COV RATIO}_i \) indicate leverage points and small values of \( \text{COV RATIO}_i \) indicate outliers.[4, p. 219]

### 3.3 Multicollinearity

When relationships between regressor are non-linear, they are said to be orthogonal. To obtain accurate results, as the relative effect of individual regressors and other estimations, conditions like orthogonality (or near orthogonality) is needed.

Opposite of regressor variables being orthogonal, is when they achieve perfect linear dependency. When this relationship instead is near-linear, the problem of multicollinearity arises. This condition leads to misleading or erroneous regression interference.

In mathematical terms, multicollinearity exists when following equation holds approximately

\[ \sum_{j=1}^{p} t_j X_j = 0 \]  

(25)

where \( t_j, \quad j = 1, ..., p \) are constants not all zero.

Severe multicollinearity may induce severe issues on the least-squares estimate of the regression coefficients. One of these being that the variances and covariances of the least-square estimators of the variables in question will be very large, meaning that data taken from same levels of \( x \) can differ widely from time to time. Another effect of multicollinearity is that it tend to create least-squares estimates too large in absolute value.[4, pp. 285-289]

#### 3.3.1 Plot of Regressor vs Regressor

In order to study pairwise correlation, regressors can be plotted against each other. If \( x_i \) and \( x_j \) create a linear relationship, it indicates that these are correlated and multicollinearity might be present. This imply that the inclusion of both regressors might not be necessary. In addition to this, the plot can detect possible influential points.[4, p. 149]
3.3.2 Correlation Matrix

The unit length scaled matrix $X'X$, called the correlation matrix, identifies pairwise multicollinearity (i.e. the dependency between two regressor variables). A value near unity (in absolute value), of the off-diagonal elements, indicates near-linear dependence.[4, pp. 292-294]

3.3.3 Variance Inflation Factors

In a multiple matter, the diagonal elements of $C = (X'X)^{-1}$ are usable for detecting multicollinearity. These diagonal elements, $C_{jj}$ may in fact be written as $C_{jj} = (1 - R^2_j)^{-1}$ where $R^2_j$ is the coefficient of multiple determination from $x_j$ regressed on the remaining p-1 regressors. If a subset of regressors obtain near linear dependency with $x_j$, the value of $R^2_j$ will be close to unity, making $C_{jj}$ very large. This gives the formula

$$VIF_j = C_{jj} = \frac{1}{1 - R^2_j}$$  \hfill (26)

Since $Var[\hat{\beta}_j] = \sigma^2 C_{jj}$, $C_{jj}$ can be interpreted as the factor increasing the variance of the regression coefficient due to multicollinearity.

If variance inflation factors exceeds 10, a coherent regression coefficient is poorly estimated on account of multicollinearity.[4, p. 296]

3.4 Quality Measures

When comparing models or evaluating the quality of one (e.g. degree of fit or linearity), different evaluation criteria and statistics can be useful to utilize.

3.4.1 Model Evaluation Criteria

Coefficient of Multiple Determination

The coefficient of multiple determination, denoted $R^2$, is a constant measuring how much of the total variability of the observations that are captured by the model

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_{Res}}{SS_T}$$  \hfill (27)

$R^2$ will increase with added numbers of regressor variables regardless of their values. A higher value of the constant is often better, but as more regressors might induce other problems, and the pace of improvement eventually decline, other evaluation criteria are good to use in combination.[4, p. 332]

Adjusted $R^2$

The adjusted $R^2$ is a modified version of $R^2$. In this measure, the number of regressor variables are taken in consideration. Here, a factor is added to counterbalance the addition of these, making the value of the constant decline at
some point (if the performance of the single added variable is not contributing enough).

\[
R_{Adj}^2 = 1 - \frac{SS_{Res}/(n - p)}{SS_T/(n - 1)}
\]  

(28)

where \(n\) is the number of observations and \(p\) is the number of variables in the model.\[4, \text{p. 333}\]

**Mean Squared Error**

The mean squared error takes the mean of the residual sum of squares (\(SS_{Res}\))

\[
MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \frac{SS_{Res}}{n}
\]  

(29)

To obtain the unbiased version of MSE, called residual mean square, one must remove the number of degrees of freedom from the denominator \[4, \text{pp. 333-334}\]

\[
RMS = \frac{1}{n - p} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \frac{SS_{Res}}{n - p}
\]  

(30)

**Mallow’s \(C_p\) Statistic**

The statistic of Mallow’s is related to the mean squared error but also takes bias in consideration

\[
C_p = \frac{SS_{Res}}{\hat{\sigma}^2} - n + 2p
\]  

(31)

If a regression model comes with zero bias, the value of \(C_p\) will be \(p\), i.e. the number of regressors used in the model. A regression model aims at a low value of Mallow’s \(C_p\).\[4, \text{pp. 334-336}\]

**Akaike information criterion**

The AIC is a log-likelihood measure, calculated as follows:

\[
AIC = n \ln (L) + 2p
\]  

(32)

where \(L\) is the likelihood function for the evaluated model. When ordinary least-squares is applied, the measure becomes

\[
AIC = n \ln \left(\frac{SS_{Res}}{n}\right) + 2p
\]  

(33)

If added regressors contribute with insufficient information, the measure will be penalized. A regression model aims at a low value of AIC.\[4, \text{p. 336}\]

**Bayesian Information Criterion**

The measure of BIC induces a great penalty for added regressor variables. BIC comes in two versions, where R uses the version of Schwartz. A regression model aims at a low value of BIC.\[4, \text{pp. 336-337}\]

\[
BIC_{Sch} = n \ln \left(\frac{SS_{Res}}{n}\right) + p \ln (n)
\]  

(34)
3.4.2 Hypothesis Testing
When coefficients are estimated, two questions arise:

1. What is the overall adequacy of the model?
2. Which regressors are of more importance?

These are answered by hypothesis testing.[4, p. 84]

3.4.2.1 Test for Significance of Regression
This procedure test the overall adequacy of the model by concluding if there is a linear relationship between the response variable $y$ and the regressors $x_1, x_2, ..., x_k$ or not. The hypothesis becomes

$$H_0 : \beta_1 = \beta_2 = ... = \beta_k = 0$$
$$H_1 : \beta_j \neq 0 \text{ for at least one } j$$

If the null hypothesis is rejected, at least one of the regressors contribute to the model.

F-statistic
The F-statistic is given by

$$F_0 = \frac{SS_R/k}{SS_{Res}/(n - k - 1)} = \frac{MS_R}{MS_{Res}}$$

and follows the $F_{k,n-k-1}$ distribution where $k$ is the number of regressors and $n-k-1$ is the degrees of freedom. If $F_0$ is large, the likelihood of some $\beta_j \neq 0$ increases. More precisely, if $F_0 > F_{\alpha,k,n-k-1}$, the null hypothesis in equation 35 should be rejected.[4, p. 86]

3.4.2.2 Tests on Individual Regression Coefficients
After a linear relationship between the response and the regressors has been stated, it is interesting to detect which of these regressors, and which do not, contribute to the linearity. To test the significance of a single coefficient, the following hypothesis test is stated

$$H_0 : \beta_j = 0$$
$$H_1 : \beta_j \neq 0$$

$t$-statistic
A statistic to decide whether to reject a specific regressor or not is the test statistic, computed as follows:

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{\sigma^2 C_{jj}}} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$
where \( C_{jj} \) is the diagonal elements of \((X'X)^{-1}\). If \( |t_0| > t_{\alpha/2,n-k-1} \) the null hypothesis of equation 37 may be rejected. Since the regression coefficient \( \hat{\beta}_j \) depend on the remaining regressor variables, this becomes a test of the contribution of \( x_j \) given all other regressors.[4, pp. 88-89]

### 3.4.3 Confidence Interval on Estimated Coefficients in Multiple Regression

In order to test the confidence of a value of a parameter, confidence intervals may be used. These rely on the assumption of independently distributed errors. If the errors hold means of zero and variances of \( \sigma^2 \), \( y_i \) is also normally and independently distributed with mean \( \beta_0+\sum_{j=1}^{k}\beta_j x_{ij} \) and variance \( \sigma^2 \). Since the estimated coefficient, \( \hat{\beta} \), is a linear combination of the regressors, \( \hat{\beta} \) is normally distributed with mean vector \( \beta \) and the covariance matrix \( \sigma^2(X'X)^{-1} = \sigma^2 C_{jj} \). This gives the t-statistic

\[
\frac{\hat{\beta}_j - \beta}{\sqrt{\sigma^2 C_{jj}}} \quad j = 0,1,\ldots,k
\]

(39)

which is t distributed with \( n-p \) degrees of freedom, where \( \hat{\sigma} \) is the estimate of the error variance. Based on these statistics it is possible to define a 100(1 - \( \alpha \)) confidence interval for coefficient \( \beta_j \) as follows:

\[
\hat{\beta}_j - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 C_{jj}} \leq \hat{\beta}_j \leq \hat{\beta}_j + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 C_{jj}}
\]

(40)

where \( se(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 C_{jj}} \) is the standard error of \( \hat{\beta}_j \).

These confidence intervals can be interpreted as boundaries of precision of the true value of a parameter. If repeated samples were taken at the same \( x \) levels and a 100(1 - \( \alpha \))% confidence interval was constructed on the estimated parameter each time, then 100(1 - \( \alpha \))% of those intervals would contain the true value of the parameter.[4, p. 98]

### 3.4.4 Cross-Validation

The procedure of cross-validation is based on dividing the observed values in test respectively training data. Training data is used to calibrate statistical methods and test data is used for validation of these.

A k-fold cross-validation divides the data into \( k \) subsets. For \( k \) iterations, each subset acts as training data. The remaining \( k-1 \) subsets constitutes the test data, different for each iteration. On this test data, several evaluation criteria are performed. When all iterations are done, a simple mean of the different calculated criteria are made.[11]
4 Data

4.1 Description of Data
Data was collected on stocks from large cap, mid cap and small cap listed on stock exchanges in Sweden, Norway, Denmark, and Finland. All the data points were collected for the last trading day of the month, between January 2014 and December 2018. The following data points have been used in the analysis in this thesis.

- Current Market Cap
- Last payed price
- Volatility 260 days
- Price to book equity ratio
- Beta raw

4.2 Data Collection
All data came from a Bloomberg terminal, in excel format. A Python script was used to transfer all data to an SQL-database for easier categorization and access. A python script was then used to extract the relevant information and create a csv-file for importing in R.

5 Methodology

5.1 Literature Study
A literature review to study mathematical regression models for equity factors was performed. Several relevant articles on the topic were found in financial journals.

The methods of this thesis are heavily influenced by the works of Fama and French[7], Carhart[3], and Black[13]. Therefore, search results have most often originated in these names. In excess of this, elaborate articles were received from Nordnet Bank, most often containing relevant research on alternative risk premia and equity risk factors. "Equity Risk Premia Strategies: Risk Factor Approach to Portfolio Management”[15] and "Systematic Strategies Across Asset Classes: Risk Factor Approach to Investing and Portfolio Management”[16] were two articles frequently used. The dialogue with Nordnet Bank and the articles provided were part of the creation of the research question of this thesis.

Lastly, the main literature used for the mathematical regression approach was Introduction to linear regression, by Douglas C. Montgomery [4].
5.2 Construction of Factors

First, the factors used in this thesis are shortly presented. Subsequent, the construction of the factors (regressors) are demonstrated. Lastly, the construction of the dependent response variable (the low market beta portfolio) is presented.

5.2.1 Description of Factors

The size factor, SMB, is a portfolio based on the differences between small cap and large cap. Size is defined as the market value of the firm considered. The value factor, HML, is a portfolio that concerns the value of the firms. The portfolio is constructed based on the differences between companies with low respectively high book-to-market ratios. The momentum factor, MOM, is a portfolio that is long on previous 12-month return winners and shorts previous 12-month loser stocks. The volatility factor, VOL, is a portfolio that is long on stock with low volatility and shorts stocks with high volatility.

5.2.2 Regressor Variables

\( R_m - R_f \)

For every month \( t \), the return of the whole universe, with all assets in the portfolio, were calculated. Risk-free rate from the four different countries were collected, based on the one-month rate on STIBOR for Sweden, EURIBOR for Finland, CIBOR for Denmark, and NIBOR for Norway. A risk-free rate for the universe was then calculated as the value-weighted mean on the inter bank rates. The regression factor is the universe market return minus the value-weighted risk-free rate.

SMB and HML

SMB and HML were constructed in similar ways. Each month, the universe was sorted based on size and two portfolios were created, separated by the median. The universe was then sorted based on value (book-to-market ratio) and divided into three portfolios, separated by the higher and lower 30th percentile. As seen in table 1 below, 6 smaller portfolios could be constructed from the intersections of the two different lists. These six portfolios were SL (small-low), SM (small-medium), SH (small-high), BL (big-low), BM (big-medium), BH (big-high). Every month, the value-weighted returns (based on the market capitalization) of these portfolios were calculated.[9]

---

*Market value is the product of a company’s shares outstanding and the market price of each share. Market value is the same as market capitalization.[14, p. 61]
<table>
<thead>
<tr>
<th>Median ME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small High</td>
</tr>
<tr>
<td>Small Medium</td>
</tr>
<tr>
<td>Small Low</td>
</tr>
</tbody>
</table>

Table 1: Construction of the SMB factor

Using the above six portfolios, the SMB factor was calculated as in equation 41.

\[
SMB = \frac{SH + SM + SL}{3} - \frac{BH + BM + BL}{3} \tag{41}
\]

The HML factor only considers the top and lower 30th percentile portfolios, and was thus calculated as in equation 42.

\[
HML = \frac{SH + BH}{2} - \frac{SL + BL}{2} \tag{42}
\]

**MOM**

The momentum factor, MOM, was compiled in a similar way as above. For each stock, the momentum was calculated as the return over a twelve months period, with the latest month’s return excluded\(^\star\).

\[
\text{momentum} = \frac{\text{Market cap 2 month ago}}{\text{Market cap 12 month ago}} \tag{43}
\]

The universe was then sorted based on the calculated momentum, and divided using the 30th and 70th percentiles. Six portfolios were constructed based on the three momentum portfolios intersected by the two size portfolios. The six portfolios based on momentum were SLm (small-low momentum), SMm (small-medium momentum), SHm (small-high momentum), BLm (big-low momentum), BMm (big-medium momentum), BHm (big-high momentum). The value-weighted returns (based on the market capitalization) were calculated and the MOM factor was constructed, for each month, as in equation 44.[10]

\[
MOM = \frac{SHm + BHm}{2} - \frac{SLm + BLm}{2} \tag{44}
\]

\(^\star\)The exclusion of the return of the twelfth month is due to the reversal effect associated with momentum. The reversal effect means that the winners of short-term momentum tend to perform poorly the next month [12].
VOL
All stocks were sorted based on the 260-days moving volatility*. Three portfolios were created using the 30th and 70th percentiles, and the value-weighted portfolio returns were calculated. With the three portfolios Vh (high volatility), Vm (medium volatility), Vi (low volatility), the VOL factor was calculated as in equation 45.

\[ VOL = Vl - Vh \] (45)

5.2.3 Response Variable
One portfolio, composed based on observed low market-beta, was used as the dependent response variable. The portfolio was constructed by sorting all stocks on the observed market-beta each month, then dividing the sorted list into ten portfolios of equal sizes. The low market beta portfolio return was subsequently calculated as the equal-weighted return of the stocks of the lower 10th percentile. The construction of the low-beta portfolio was based on the work of Black (1972), who used a similar portfolio construction method to test the CAPM model.[13] The beta of a stock was received as part of the data set from the Bloomberg terminal, and is described as the observed volatility of the price of a stock compared to the volatility in the market.

5.3 Model for Testing
The regression analysis was performed using RStudio with the following model

\[ E[R_i] = R_f + \beta_{m,i} (E[R_m] - R_f) + \beta_{s,i}SMB + \beta_{v,i}HML + \beta_{m,i}MOM + \beta_{vol,i}VOL + \alpha \] (46)

6 Results
In the section of results, different plots and tables are presented. The main output is emphasized, but further discussed in the discussion section.

6.1 Financial Returns
In figure 1, the cumulative returns of both the response and each regressor are presented. The cumulative return for the 60th month is presented in table 2

---

*A measure of the risk of price movements for a security, calculated from the standard deviation of day to day logarithmic historical price changes. The 260-day price volatility equals the annualized standard deviation of the relative price change for the 260 most recent trading days closing price, expressed as a percentage.[14, p. 356]
Table 2: Return for the 60th month

<table>
<thead>
<tr>
<th>Low-Beta</th>
<th>( R_m - R_f )</th>
<th>SMB</th>
<th>HML</th>
<th>MOM</th>
<th>VOL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Return:</td>
<td>2.201895</td>
<td>1.206051</td>
<td>0.6204501</td>
<td>0.2888444</td>
<td>1.899765</td>
</tr>
</tbody>
</table>

6.2 Multiple Linear Regression Analysis

6.2.1 Residual Analysis

Quantile-Quantile Plots
The quantile-quantile plot (Q-Q plots) presented in figure 2 shows that the basic data of the thesis follow a normal distribution, even if marginal signs of a heavy-tailed distribution are present. Since the multiple linear model assumes normality in the underlying data, no transformation is needed. In excess of this, the plot denotes observation 1 a possible influential point.
Scaled Plots
Below is the standardized and studentized residuals presented. These plots indicates that data point 1 must be further investigated, due to the fact that this point is located outside the permitted region $d_1 > 3$ (stated by Montgomery [4, p. 131]). The increase of this value in the studentized residual plot indicates some degree of leverage, since the diagonal elements of the hat matrix is a part of equation 14.
Plot of Residuals vs Fitted Values
Figure 4 shows a horizontal pattern, i.e. the variance of the errors possesses homoscedasticity. Observation 1 is deviating, possibly indicating an outlier.

Figure 4: Residuals vs fitted values

Plot of Residuals vs Regressor
Overall, all observations of the plots in figure 5 are captured by horizontal boundaries.

Figure 5: R-Student residuals plotted against each regressor
Partial Residual Plots
Figure 6 shows that the market factor, the value factor and the momentum factor contribute with poor information to the model. The remaining factors, i.e the volatility factor and the size factor, are valuable to the model.

![Added-Variable Plots](image)

Figure 6: Added-variable plots

### 6.2.2 Influence and Leverage

**Hat Matrix**

Figure 7 maps out the values of the diagonal elements of the hat matrix. Values above the horizontal line, located at the cutoff value $2p/n$, warrants attention. Here, $n$ is the number of observations and $p$ is the number of coefficient estimates.
Figure 7: Diagonal elements of the hat matrix

As can be read in table 3, two values are located outside the permitted area. These observations are thereby classified as leverage points.

<table>
<thead>
<tr>
<th>Index</th>
<th>$h_{ii}$</th>
<th>p</th>
<th>n</th>
<th>Cutoff Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>0.24031371</td>
<td>6</td>
<td>60</td>
<td>0.2</td>
</tr>
<tr>
<td>59</td>
<td>0.26926967</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Deviating figures of diagonal elements of the hat matrix

**COVRATIO**

Figure 8 displays the values of $COVRATIO_i$ that are enclosed by cutoff values (and those who are not).
Observations 1 and 15 meet characteristics of an outlier, i.e. they fall below the cutoff value $1 - 3 \times p/n$. The remaining data that fall out of bounds, i.e. exceed $1 + 3 \times p/n$, indicate leverage. These observations are presented in Table 4.

<table>
<thead>
<tr>
<th>Index</th>
<th>$COVRATIO_i$</th>
<th>p</th>
<th>n</th>
<th>Cutoff Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.3389244</td>
<td>6</td>
<td>60</td>
<td>$1 \pm 0.3$</td>
</tr>
<tr>
<td>15</td>
<td>0.5925237</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>1.3620696</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>1.3306552</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>1.4361032</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>1.3113650</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>1.3340869</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>1.3688888</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Deviating observations of $COVRATIO_i$

Observation 1 holds by far the most severe value. Despite being notable, the rest of the data are somewhat stable (i.e. values are close to the cutoff values).

**DFFITS**

Figure 9 shows that quite few data points fall outside the cutoff boundaries set by $\pm 2 \sqrt{p/n}$.
Table 5 state observations that might be influential.

<table>
<thead>
<tr>
<th>Index</th>
<th>$DFFITS_i$</th>
<th>p</th>
<th>n</th>
<th>Cutoff Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.7860106026</td>
<td>6</td>
<td>60</td>
<td>$\pm0.632$</td>
</tr>
<tr>
<td>14</td>
<td>0.7853848750</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.7960364383</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Deviating figures of diagonal elements of $DFFITS_i$

Observation 14 and 20 hold noticeable values, but not severe. Data point 1 stands out remarkably, being far greater than its descendant values.

**Cook’s Distance**

For the low-beta portfolio, figure 10 denotes points 1, 14 and 20 as the most influential. However, not one of these exceeds a value of 1 (the cutoff value for Cook’s distance). Even if data point 1 holds the highest value, it still falls below 0.5.
Figure 10: Cook’s distance

Figure 11: Residuals against leverage
6.2.3 Adjustments for Influence

Based on above results, observation 1 was removed. As seen in figure 13, figure 14 and figure 15 in Appendix A, assumptions about normality and homoscedasticity were not violated.

6.2.4 Multicollinearity

Regressor Plots

Plots are presented on VOL vs \( R_m - R_f \), MOM vs \( R_m - R_f \) and VOL vs MOM. All of these show vague signs of linear relationships.

Figure 12: Regressor \( i \) against regressor \( j \)

Correlation Matrix

The correlation matrix below shows small pairwise multicollinearity. SMB and HML obtain the lowest correlation, followed by the correlation between MOM respectively VOL and the market. Also, the correlation between MOM and VOL was low.
Table 6: Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>$R_m - R_f$</th>
<th>SMB</th>
<th>HML</th>
<th>MOM</th>
<th>VOL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_m - R_f$</td>
<td>1.00000000</td>
<td>-0.33472113</td>
<td>-0.18436717</td>
<td>0.03954482</td>
<td>-0.09534877</td>
</tr>
<tr>
<td>SMB</td>
<td>-0.33472113</td>
<td>1.00000000</td>
<td>0.02455941</td>
<td>0.24210913</td>
<td>-0.27452770</td>
</tr>
<tr>
<td>HML</td>
<td>-0.18436717</td>
<td>0.02455941</td>
<td>1.00000000</td>
<td>-0.45144936</td>
<td>-0.13551140</td>
</tr>
<tr>
<td>MOM</td>
<td>0.03954482</td>
<td>0.24210913</td>
<td>-0.45144936</td>
<td>1.00000000</td>
<td>0.13182968</td>
</tr>
<tr>
<td>VOL</td>
<td>-0.09534877</td>
<td>-0.27452770</td>
<td>-0.13551140</td>
<td>0.13182968</td>
<td>1.00000000</td>
</tr>
</tbody>
</table>

Variance Inflation Factor

Presented in the table below are the variance inflation factors. All regressors obtained low values of $VIF$s relative a cutoff value of 10, where the lowest number possible is 1.

<table>
<thead>
<tr>
<th></th>
<th>$R_m - R_f$</th>
<th>SMB</th>
<th>HML</th>
<th>MOM</th>
<th>VOL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$VIF_i$</td>
<td>1.248366</td>
<td>1.427622</td>
<td>1.321297</td>
<td>1.420027</td>
<td>1.210442</td>
</tr>
</tbody>
</table>

Table 7: Variance inflation factors

6.2.5 Model Characteristics

Table 8 presents all values of coefficients for the low-beta portfolio. Notice that observation 1 was removed from the regressed sample.

<table>
<thead>
<tr>
<th></th>
<th>$R_m - R_f$</th>
<th>SMB</th>
<th>HML</th>
<th>MOM</th>
<th>VOL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.02266</td>
<td>-0.06084</td>
<td>1.04541</td>
<td>0.11553</td>
<td>0.02717</td>
</tr>
</tbody>
</table>

Table 8: Estimated coefficients of model

The excess return of the low-beta portfolio shows the highest correlation to the size factor. The low volatility coefficient is the second largest closely followed by the value coefficient. The role of the market factor as well as the momentum factor is diminished.

The intercept from the model in equation 46 consists of both $\alpha$ and $R_f$. Thus, $\alpha = Intercept - R_f$, and the result is presented in table 9.

<table>
<thead>
<tr>
<th></th>
<th>Intercept:</th>
<th>0.02266</th>
<th>$\bar{r}_f$:</th>
<th>0.00109</th>
<th>$\alpha$:</th>
<th>0.02157</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 9: Alpha

30
**F-test and t-test**

As seen in table 10, the value of the F-statistics exceeds $F_{a,k,n-k-1}$ by far - for all levels of $\alpha$. The removal of null hypothesis indicates that the response variable holds a linear relationship with at least one of the regressors.

Furthermore, t-statistics indicate that regressors $R_m - R_f$, HML and MOM answer the null hypothesis, i.e. they do not hold a linear relationship with the response.

| Original | $t_0$ | $\alpha$ | $P(F \leq f)$ | $|t_{0.05/2,54}|$ |
|----------|------|--------|-------------|----------------|
| (Intercept) | 6.355 | 0.05 | 0.95 | 2.004879 |
| $R_m - R_f$ | -0.360 | 0.01 | 0.99 | 2.669985 |
| SMB | 7.851 | 0.001 | 0.999 | 3.480016 |
| HML | 0.323 | | | |
| MOM | -0.724 | | | |
| VOL | 2.397 | | | |
| **Model** | **15.66** | **0.05** | **0.95** | **2.38607** |
| | | **0.01** | **0.99** | **3.376912** |
| | | **0.001** | **0.999** | **4.836359** |

| Improved | $t_0$ | $\alpha$ | $P(F \leq f)$ | $|t_{0.01/2,53}|$ |
|----------|------|--------|-------------|----------------|
| (Intercept) | 6.200 | 0.05 | 0.95 | 2.005746 |
| $R_m - R_f$ | -0.687 | 0.01 | 0.99 | 2.671823 |
| SMB | 6.717 | 0.001 | 0.999 | 3.483777 |
| HML | 0.891 | | | |
| MOM | 0.252 | | | |
| VOL | 2.043 | | | |
| **Model** | **13.37** | **0.05** | **0.95** | **2.389444** |
| | | **0.01** | **0.99** | **3.38414** |
| | | **0.001** | **0.999** | **4.851596** |

**Table 10: F-test and t-test**

**Confidence Intervals on Regression Coefficients**

The confidence intervals of the intercept and the size factor leave out the value of zero for all significance levels. Considering the volatility factor, this only holds for a significance level of $\alpha = 0.05$. 

31
<table>
<thead>
<tr>
<th>$\alpha = 0.05$</th>
<th>Original</th>
<th>Improved</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>(0.01735097, 0.03334415)</td>
<td>(0.01532733, 0.02998685)</td>
</tr>
<tr>
<td>$R_m - R_f$</td>
<td>(-0.23259018, 0.16180709)</td>
<td>(-0.23857316, 0.1168763)</td>
</tr>
<tr>
<td>SMB</td>
<td>(0.93986563, 1.58445570)</td>
<td>(0.73325634, 1.3575323)</td>
</tr>
<tr>
<td>HML</td>
<td>(-0.24020850, 0.33260504)</td>
<td>(-0.14452022, 0.37557718)</td>
</tr>
<tr>
<td>MOM</td>
<td>(-0.31483049, 0.14782651)</td>
<td>(-0.18904417, 0.24338046)</td>
</tr>
<tr>
<td>VOL</td>
<td>(0.03047132, 0.34168696)</td>
<td>(0.00265954, 0.28590385)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\alpha = 0.01$</th>
<th>Original</th>
<th>Improved</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>(0.01469816, 0.03599696)</td>
<td>(0.01289323, 0.03242095)</td>
</tr>
<tr>
<td>$R_m - R_f$</td>
<td>(-0.29800953, 0.22722643)</td>
<td>(-0.29759463, 0.17590909)</td>
</tr>
<tr>
<td>SMB</td>
<td>(0.83294637, 1.69137496)</td>
<td>(0.62959345, 1.46123613)</td>
</tr>
<tr>
<td>HML</td>
<td>(-0.33522206, 0.42761860)</td>
<td>(-0.23087830, 0.46193525)</td>
</tr>
<tr>
<td>MOM</td>
<td>(-0.39157219, 0.22456822)</td>
<td>(-0.26084488, 0.31518117)</td>
</tr>
<tr>
<td>VOL</td>
<td>(-0.02115055, 0.39330883)</td>
<td>(-0.04437094, 0.33293433)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\alpha = 0.001$</th>
<th>Original</th>
<th>Improved</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>(0.01146730, 0.03922782)</td>
<td>(0.00992604, 0.03538815)</td>
</tr>
<tr>
<td>$R_m - R_f$</td>
<td>(-0.37768368, 0.30690059)</td>
<td>(-0.36954244, 0.24785690)</td>
</tr>
<tr>
<td>SMB</td>
<td>(0.70272953, 1.82159181)</td>
<td>(0.50322725, 1.58760232)</td>
</tr>
<tr>
<td>HML</td>
<td>(-0.45093897, 0.54333551)</td>
<td>(-0.33614973, 0.56720668)</td>
</tr>
<tr>
<td>MOM</td>
<td>(-0.48503583, 0.31803186)</td>
<td>(-0.34837072, 0.40270700)</td>
</tr>
<tr>
<td>VOL</td>
<td>(-0.08402077, 0.45617904)</td>
<td>(-0.10170162, 0.39026500)</td>
</tr>
</tbody>
</table>

Table 11: Confidence intervals on regression coefficients
### Summary of Statistics

|          | Estimate | Std. Error | t-value | Pr(>|t|)          |
|----------|----------|------------|---------|------------------|
| (Intercept) | 0.025348 | 0.003989   | 6.355   | 4.57 · 10^{-8}  *** |
| $R_m - R_f$ | -0.035392 | 0.098359   | -0.360  | 0.720            |
| SMB      | 1.262161 | 0.160755   | 7.851   | 1.71 · 10^{-10} *** |
| HML      | 0.046198 | 0.142855   | 0.323   | 0.748            |
| MOM      | -0.083502 | 0.115383   | -0.724  | 0.472            |
| VOL      | 0.186079 | 0.077615   | 2.397   | 0.020            *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '

F-statistic: 15.66  
p-value: 1.654 · 10^{-9}  
Degrees of freedom: $DF_1 = 5$, $DF_2 = 54$  
Residual std. error: 0.02107  
Multiple $R^2$: 0.5919  
Adjusted $R^2$: 0.5541  

Table 12: Summary of statistics for full model

|          | Estimate | Std. Error | t-value | Pr(>|t|)          |
|----------|----------|------------|---------|------------------|
| (Intercept) | 0.022657 | 0.003654   | 6.200   | 8.69 · 10^{-8}  *** |
| $R_m - R_f$ | -0.060843 | 0.088611   | -0.687  | 0.495            |
| SMB      | 1.045415 | 0.155632   | 6.717   | 1.29 · 10^{-8}  *** |
| HML      | 0.115528 | 0.129652   | 0.891   | 0.377            |
| MOM      | 0.027168 | 0.107796   | 0.252   | 0.802            |
| VOL      | 0.144282 | 0.070608   | 2.043   | 0.046            *|

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '

F-statistic: 13.37  
p-value: 1.941 · 10^{-8}  
Degrees of freedom: $DF_1 = 5$, $DF_2 = 53$  
Residual std. error: 0.01893  
Multiple $R^2$: 0.5578  
Adjusted $R^2$: 0.5161  

Table 13: Summary of statistics for improved model
Model Evaluation Criteria
Below are the evaluation criteria, for the full model and the improved model, presented:

<table>
<thead>
<tr>
<th>Model</th>
<th>Adjusted $R^2$</th>
<th>Mallow’s $C_p$</th>
<th>AIC</th>
<th>BIC</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>0.5541</td>
<td>6.000000</td>
<td>-285.2255</td>
<td>-270.5651</td>
<td>0.0003996311</td>
</tr>
<tr>
<td>Improved</td>
<td>0.5161</td>
<td>6.000000</td>
<td>-293.0164</td>
<td>-278.4736</td>
<td>0.0003218115</td>
</tr>
</tbody>
</table>

Table 14: Model evaluation criteria

Cross-Validation
Cross-Validation was made with ten folds, for both the full and the improved model.

<table>
<thead>
<tr>
<th>Model</th>
<th>Adjusted $R^2$</th>
<th>Mallow’s $C_p$</th>
<th>AIC</th>
<th>BIC</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>0.5284416</td>
<td>4.733296</td>
<td>-257.0603</td>
<td>-251.0934</td>
<td>0.0004872074</td>
</tr>
<tr>
<td>Improved</td>
<td>0.4930861</td>
<td>4.361363</td>
<td>-264.3548</td>
<td>-258.4383</td>
<td>0.0003870218</td>
</tr>
</tbody>
</table>

Table 15: Model evaluation using cross-validation

7 Discussion
In this thesis, a multiple ordinary least regression on a portfolio of Nordic stocks with low market beta, regressed on different risk factors, has been performed. The model analyzed was Carhart’s four-factor model extended with a volatility factor. The research question stated was:

*How does a set of risk factors correlate with the return of a low market beta portfolio?*

The results show that half of the variation of the low-beta portfolio can be explained by the set of factors used ($R_m - R_f$, Vol, Mom, HML, MOM). This finding unable further conclusions about if alpha is captured by the present set of risk factors [16]. Especially, it leaves room for improvement of the model, either by a different set of risk factors or by the inclusion of more factors.

The low-beta portfolio is highly correlated with the SMB-factor. Furthermore, the portfolio is zero-correlated with the market beta, which mainly is due to the design and construction of the low-beta portfolio using actual observed past values. The portfolio correlated positively with HML and VOL, but only to a small extent. Furthermore, the portfolio is almost uncorrelated to the MOM-factor.
7.1 Limitations and Method Critique

The method of multiple linear regression is a successful and commonly used approach in previous research of factor analysis. The construction of the factors in Carhart’s model is also motivated with evidence. However, since the volatility factor is of later research, less evidence on the construction of this factor exist (implying greater uncertainty). In excess of this, the derivation of the dependent variable, based on the theory of Black, has a different approach from the rest of the value-weighted portfolios (i.e. the regressors) since it uses an equal-weighted portfolio [13]. Effects of combining the two approaches is not presented in this thesis. However, as stated in the working paper ”Equal or Value Weighting? Implications for Asset-Pricing Tests”[18], an equal-weighted portfolio has higher exposure to equity risk factors than value-weighted.

The lower limit of stocks included is based on the specific countries’ small cap indexation, which might be different between countries. This can result in the exclusion of smaller stocks in one country, but the inclusion of stocks with the same capitalization in another country. Since the asset is listed on the small, medium or large cap, it is however assumed that the listing in itself increases the amount of trading. There is, nevertheless, still a risk that some of the small cap stocks have low trading activity, jeopardizing the true market valuation of the stocks.

The model that was chosen explains about half of the variation in the returns. However, this performance was only linked to the response portfolio tested (i.e. the low market beta portfolio) - general conclusions about the factors would be more accurate if more portfolios were tested.

7.2 Mathematical Approach

Homoscedasticity was justified by figure 4 and figure 5 where observations were contained by a horizontal band, indicating constant variance. Observation 1 was deviating, showing sign of an outlier.

When a constant variance of the errors was confirmed, a test for normality was done. Montgomery (2012) states that a larger sample size $n \geq 32$ produces accurate results for Q-Q-plot - this thesis used 60 in the regression analysis [4, p. 139]. Since there was rather many data points, the model was considered robust. Even if the plot was slightly skewed, overall linearity in the Q-Q-plot made the assumption of normality hold. However, deviating points indicated that further methods should be used in order to specify the data point as influential or not. As seen in figure 2, observation 1 had to be examined.

Potential outliers were assessed in the residual analysis. Initially, in figure 3 observation 1 was outside the boundary at level 3. This upper bound is stated by Montgomery (2012), though intended for the standardized residual. As seen in the plot, the studentized residuals enhanced the deviation of observation
1. This is an indication of leverage since the formula for studentized residuals partly consists of the diagonal elements of the hat matrix.[4, p.130]

Figure 7 presents the diagonal elements of the hat matrix, indicating leverage points. Observation 59 obtained the highest value, located outside the boundary for cutoff values \((2p/n)\), and was thus highlighted as a leverage point. However, by looking at the residual plots, observation 59 was not deviant, indicating that the observation was not influential. The plot also showed that observation 1 was located close to the boundary for cutoff values, even if it existed in the permitted area.

To effectively indicate influential points, diagnostics that combined analysis of leverage and outliers were utilized. Cook’s distance informed three points: 1, 14 and 20. Even if the value of observation 1 was over four times as big as the second largest, it did not obtain a value over 0.5 (far below the cutoff value of 1). However, studying the values of \(COVRATIO_i\) and \(DFITS_i\), observation 1 was remarkably distant from the cutoff values. Based on this, observation 1 was removed from the sample.

As seen in table 4, several values of \(COVRATIO_i\) exceeded or fell below the cutoff values. However, since most of these were close to the boundary for permission and were not notified in other influence measures, these data points were left untouched.

To examine multicollinearity, the correlation matrix and variance inflation factors were performed and examined. The off-diagonal elements of the correlation matrix showed low pairwise correlation, especially between the SMB and HML factors. Also, the correlation between VOL and MOM was low, as for VOL and MOM against the market factor \((R_m - R_f)\) respectively.

Conclusions about factor correlation could have been deceptive if the correlation between each factor and the rest of the factors was not examined (i.e. multicollinearity). The variance inflating factors covered this and were in fact low, indicating minimal multicollinearity. This showed that the information provided by each factor coefficient was addressed to the factor examined - it was not a result of the behavior of the remaining factors.

One of many benefits with low multicollinearity is the fact that the information given by the partial regression plots are not misinformative. For this thesis, it means that the added-variable plots show the correct relationship between the response and the regressor variables.[4, p. 145]

To further validate results from the added-variable plots, t-statistics as well as F-statistics were performed. Table 10 shows results for both the full model and the improved model. F-statistics confirmed that the models satisfied a linear relationship against the response variable. Results of t-statistics did however
For both the original and the improved model, only some of the regressors showed signs of statistical significance, i.e., the null hypothesis could be dismissed and a linear relationship could be proven. The factors referred to are SMB for all significance levels and the volatility factor for a significance level of 0.05. When the influential point was removed, all t-statistics were worsened except for the ones coherent to the value factor (HML) and the momentum factor (MOM). Despite this, the volatility factor (VOL) still showed statistical significance at level 0.05. However, it was now closer to the upper bound of the $t_{\alpha/2.53}$ distribution.

Statistical significance was achieved for the intercept in both models, meaning that a zero intercept could be rejected. This was in line with the used model in equation 46, where the risk-free rate is a part of the intercept.

Confidence intervals were as well performed at different significance levels. If the span of an interval leaves out the value of zero, there is good evidence that the estimated coefficient contributes to the model (if zero is included, the opposite prevails).

In accordance with the t-statistics, the SMB factor was emphasized the most in showing a linear relationship with the response variable. At a confidence level of 95%, the volatility factor also obtained a confidence interval beyond zero. In excess of this, an intercept of zero could be excluded.

The rest of the factors held intervals that crossed the value of zero. Out of these, the momentum factor and the market factor had the smallest span surrounding zero, indicating that these factors, with greatest certainty, contributed poorly to the model. As observation 1 was removed, the lengths of the confidence intervals were reduced for all significance levels, indicating that estimation confidence was increased.

Evaluation criteria chosen were Adjusted $R^2$, Mallow’s $C_p$, AIC, BIC and MSE. When expanding a model with additional factors, regular $R^2$ will regardless increase. Since the used model was an expansion of older models, adjusted $R^2$ was used to penalize for additional factors that did not contribute with additional information.

All criteria, except for the adjusted $R^2$, were improved when observation 1 was excluded from the data set. As the adjusted $R^2$ was lowered, the removal of the influential point could first seem contradictory as the intention was to improve the model. However, the thesis aims at describing certain factors’ significance, not to manipulate these in order for improvement. Observation 1 did not reflect assumptions about normality, as other observations did, and was thereby removed.

To verify the values and the changes in these, cross-validation was performed.
Values were lowered in comparison with the values of the evaluation criteria, but the order between these remained. The number of folds was considered appropriate since the used data was sufficient. Furthermore, a number of 10 folds represented the variation of the underlying distribution in the data.

### 7.3 Aspects of Industrial Economics

After the 2008 financial crisis, methods of equity risk factor investing became increasingly popular among investors for two reasons: To decrease portfolio volatility, and to still receive premia for being invested in risk factors [15, p. 3]. The investor sentiment of decreasing portfolio volatility makes this thesis’ low-beta portfolio especially interesting, since the return had a positive correlation with the volatility risk factor (VOL).

The correlation between factors in this thesis can be compared with results from Fama and French’s research study on American stocks and bonds between 1963-1991. In the study, HML to SMB obtained a correlation value of 0.08, compared to this thesis’ value of 0.02. The correlation of SMB to $R_m - R_f$ was 0.32, compared to this thesis’ value of -0.33. Lastly, the correlation of HML to $R_m - R_f$ was -0.38, compared to this thesis’ value of -0.18. Although this thesis and the study of Fama and French are separated by more than twenty years and performed in different countries, the similar correlation values indicate somewhat similar factor conditions now as then.[7]

The low correlation between factors presented in this thesis and the correlation of the low-beta portfolio with some of the risk factors, corresponds to the theory about ARP. To get a premia based on a specific factor that is zero-correlated with the market, is the essence of alternative risk premia.

Over the five years studied, an investment in this portfolio strategy would have yielded a return factor of 2.2, while being very low correlated with underlying market exposure. If the first month (i.e observation 1), considered an influential point, was excluded, the return factor would have decreased to 1.93. This contributes to the fact that the removal of the data point was accurate, since the inclusion would increase returns by 27% - compared with the other stock performances, a relatively unusual and high percentage growth.

The portfolio return can be compared with the market performance which yielded a return factor of 1.289 for the full five year period, and with the single SMB-factor which yielded a return factor of 0.62. It may seem counter intuitive that the highest correlation was between the SMB-factor and the low-beta portfolio, while the SMB-factor portfolio produced a return factor less than 1. This is simply explained by the fact that the returns of the both portfolios followed similar patterns, but possessed different mean values.
The second largest factor that the low-beta portfolio correlated to was the volatility factor (VOL). This portfolio produced a factor return of 0.993. As such, the portfolio seems to value small businesses higher than large businesses (due to the SMB factor), while at the same time valuing small volatility over high volatility. This could be described as investing in small cap stocks that behave like large cap stocks (lower volatility), and shorting the opposite. The momentum factor obtained the lowest correlation with the low-beta portfolio, which indicated that the stocks in the portfolio was not among the winners or losers, but rather in between.

Based on the results, an investor could make use of the low-beta portfolio strategy to get a leverage on the SMB-factor, while at the same time being unexposed to the market and heavier invested in low volatile stocks.

### 7.3.1 Factor Classification

In order to suggest additional, possibly contributing factors to the model, this thesis’ choice of equity risk factors can be compared with research done by financial analysts at JPMorgan. The investment banking institution separate alternative risk factors into five broad categories: Value, Growth, Quality, Momentum, and Volatility.

Value includes factors that measure if the stock appears cheap or expensive. These risk factors include earnings yield, free cash flow yield, and dividend yield. Growth includes factors that consider companies likely to deliver strong earnings growth, and include measures such as earnings momentum, price/earnings to growth, and free cash flow to invested capital. Quality includes factors that measure the profitability and how sustainable such profitability is. This includes measures such as profit margin, asset turnover, leverage and other return on equity-factors. Momentum measures the technical aspects of time series analysis, how the price of the stock changes over time. Volatility includes factors that measure the deviation of stock returns over different time periods.[15]

Compared to above classification of equity risk factors, this thesis HML factor would belong to the Value-classification and the SMB factor would belong to the Volatility-classification*. The volatility factor and the momentum factor speak for themselves. Using this model of classification, including factors with measures on the Quality-classification and the Growth-classification could possibly have contributed with describing the return of the low-beta portfolio.

---

*According to analysts of JPMorgan, SMB belongs to the Volatility-classification due to close linkage between volatility premia and small size premia.[15, p. 28]
7.4 Further Research

In future studies, it would be interesting to add more alternative risk premia factors to analyze the low-beta portfolio. It would also be interesting to analyze portfolios based on low volatility and momentum, and regress on a set of similar factors as in this thesis.

8 Conclusion

The aim of the thesis was to perform a multiple linear regression analysis of a low-beta portfolio in order to study its correlation with given data and constructed factors. In terms of economics, it was investigated in how the returns of the portfolio could be described by risk factors. Two of these risk factors, in some literature denoted as alternative risk premia, were momentum and low volatility.

Significantly low correlation was shown between the regressors, both pairwise and multidimensional. Rather than showing near-linear dependencies with each other, the regressors of the thesis approached orthogonality. This was in full alignment with the theory of alternative risk premia.

Mathematically, this meant that the relative effects of a regressor variable could be identified without interference of the remaining regressors. Further, low multicollinearity justified the results of other methods used in this thesis. In terms of economics, these results implied evidence for risk diversification. However, presence of causality can never be determined simply by multiple linear regression.

Returns of the low-beta portfolio had the strongest correlation with the size factor (SMB), followed by the volatility factor (VOL). Relative to SMB, VOL was however very low. Yet, compared with the market factor, the volatility factor performed well. Since only the market factor was manipulated to correlate poorly with the low-beta portfolio, it was interesting to see that the volatility factor exhibited relevancy in describing the returns.

As for the momentum coefficient, results were different. This alternative risk premia contributed not much more than the market factor in describing the excess return of the portfolio. However, as the momentum factor portfolio in itself performed best regarding excess return, theory about the ability of its characteristics to capture equity was confirmed.

The model capture about half of all variation of returns. This made it hard to conclude to what extent the different betas contributed in reducing the value of alpha. The shortcomings of the model opens up for improvements, such as re-specification of risk premia or the addition of others.
To sum up, constructing a low-beta portfolio and analyze what factors describe the excess returns is interesting for several reasons. Diversification has always been valuable to investors, as well as receiving equity premia that possesses low correlation with the market conditions. This thesis contributes to the fact that presently used models might have reasons to be expanded.
9 Appendix A

Figure 13: QQ-plot on standardized residuals (improved model)

Figure 14: Residuals vs fitted values (improved model)
Figure 15: R-Student residuals plotted against each regressor (improved model)

Figure 16: Added-variable plots (improved model)
Figure 17: Residuals against leverage (improved model)

Figure 18: Cook's distance (improved model)
\section{Appendix B}

\textbf{General Model of Multiple Linear Regression}

Regression evaluation is done to analyze the relation between the dependent variable, i.e. the response variable, and the independent variable, the regressor. When several independent variables are considered and the relationship between these and the response are linear (i.e. the coefficients are not exponential), multiple linear regression can be used. With \( k \) regressors, a general model can be written as follows:

\[ y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \epsilon_i \quad (47) \]

where \( k \) is the number of regressors, \( i = 1, \ldots, n \) is the observation index and \( \epsilon_i \) is the random error of observation \( i \).

The above model can also be written on a more convenient form with matrix notation, as follows:

\[ y = X\beta + \epsilon \quad (48) \]

where

\[ y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \]

and \( n \) is the number of observations.

The coefficients \( \beta \) are on beforehand unknown, and are to be selected with ordinary least-squares.\[4, \text{pp. 68-70}\]

\textbf{Assumptions About Normality}

For equations 47 and 48 to deliver accurate results, five assumptions must be fulfilled:

1. The response and the regressors hold a linear relationship, at least approximately.
2. The mean of the error term is zero, i.e. \( E[\epsilon_i] = 0 \).
3. The variance of the error term is constant, i.e. \( \text{Var}[\epsilon_i = \sigma^2] \).
4. The errors of the model is uncorrelated, i.e.

\[ \text{Cov}(\epsilon_i, \epsilon_j) = \begin{cases} 0 & \text{if } i \neq j \\ \text{Var}[\epsilon_i] & \text{if } i = j \end{cases} \]
5. The errors are normally distributed

The above assumption constitute that the regression model has a mean of \( \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} \) and a variance of \( \sigma^2 \). If these assumptions are violated, assumptions may lead to misleading results.

Together, point 4 and 5 constitutes the fact that the errors are independent random errors, where the latter assumption is necessary for hypothesis testing and interval estimation.[4, p. 129]

**Heteroscedasticity**

Heteroscedasticity is the phenomenon that arises when the third normality statement is violated. This is often due to model errors following a distribution other than the normal, where the variance depends on the value of the dependent variable.

If not considered and adjusted for, this might harm the interpretation of the model performance. The opposite of heteroscedasticity is homoscedasticity.[4, p. 170]

**Ordinary Least-Squares**

One way of establishing the relationship between the response and regressors, is by compiling the method of ordinary least-squares. This is performed by choosing the best linear unbiased estimator (\( \hat{\beta} \)) such that the squared error term is minimized. In other words, minimizing the following expression:

\[
S(\beta) = \sum_{i=1}^{k} \epsilon_i^2 = \epsilon' \epsilon = (y - X\beta)'(y - X\beta)
\]  
(49)

By optimizing this (through derivation and setting to zero), the least-squares normal equations are yielded:

\[
X'X\hat{\beta} = X'y
\]

and can, through inversion of the matrix \( X'X \), be rewritten as:

\[
\hat{\beta} = (X'X)^{-1}X'y
\]

(51)

For this to hold, the inverse of \( X'X \) must exist, which will always occur if the regressors are linearly independent. Furthermore, the theorem of Gauss-Markov claims that \( \hat{\beta} \) is the best linear unbiased estimator (\( E[\hat{\beta}] = \beta \)) with the smallest variance (\( Var[\hat{\beta}] = \sigma^2(X'X)^{-1} \)).[4, p. 80]

With the optimal vector of beta coefficients present, the fitted general regression model can be written as

\[
\hat{y} = X\hat{\beta}
\]

(52)
Consequently, the residuals, i.e. the distances from the real observations to the fitted line, may be minimized by by the following equation:[4, pp. 70-73]

\[ e = y - \hat{y} \]  

(53)

**Analysis of Variance (ANOVA)**

Total variance of variables is stated as follows:

\[ SS_T = y_i - \bar{y} \]  

(54)

where \( y_i \) is the \( i \)th response variable and \( \bar{y} \) is the mean of all of the response variables.

In the setting of ANOVA, the total variances are in turn partitioned in two sources of variation:

**Model Sum of Squares**

The model sum of squares (\( SS_R \)) describe to what extent the fitted model can describe the variation in the observed values. It is written as

\[ SS_R = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \]  

(55)

To obtain an unbiased estimate of \( SS_R \) in multiple regression, one must divide \( SS_R \) by the degrees of freedom as below.[4, pp. 26,85]

\[ MS_R = \frac{SS_R}{k} \]  

(56)

where \( k \) is the number of regressors in the model.

**Residual Sum of Squares**

The residual sum of squares (\( SS_{Res} \)) is a measure of unexplained deviation between the observed values and the fitted model. The smaller the value of \( SS_{Res} \) is, the better the model is fitted.

\[ SS_{Res} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]  

(57)

Below is the unbiased estimate of \( SS_{Res} \) in multiple regression presented.[4, pp. 21,85]

\[ MS_{Res} = \frac{SS_{Res}}{n - p} \]  

(58)

where \( n \) is the number of observations and \( p \) is the number of estimated parameters in the model, i.e. \( p = k + 1 \).

The total variation then becomes

\[ SS_T = SS_{Res} + SS_R \]  

(59)
The different measures of variation is often presented in the ANOVA table, stated below.[4, p. 85]

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>$F_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>$SS_R$</td>
<td>$k$</td>
<td>$MS_R$</td>
<td>$MS_R/MS_{Res}$</td>
</tr>
<tr>
<td>Residual</td>
<td>$SS_{Res}$</td>
<td>$n - k - 1$</td>
<td>$MS_{Res}$</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>$SS_T$</td>
<td>$n - 1$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 16: ANOVA in Multiple Regression
References


