



**ROYAL INSTITUTE
OF TECHNOLOGY**

MASTER THESIS

Smart Beta Investment Based on Macroeconomic Indicators

DEPARTMENT OF MATHEMATICS,
ROYAL INSTITUTE OF TECHNOLOGY (KTH)

Author:
Alexandra Andersson

Supervisors:
Boualem Djehiche, KTH
Elena Westerdahl, Nordea

September 20, 2015

Abstract

This thesis examines the possibility to find a relationship between the Nasdaq Nordea Smart Beta Indices and a series of macroeconomic indicators. This relationship will be used as a signal-value and implemented in a portfolio consisting of all six smart beta indices. To investigate the impact of the signal-value on the portfolio performance, three portfolio strategies are examined with the equally weighted portfolio as a benchmark. The portfolio weights will be re-evaluated monthly and the portfolios examined are the mean-variance portfolio, the mean-variance portfolio based on the signal-value and the equally weighted portfolio based on the signal-value.

In order to forecast the performance of the portfolio, a multivariate GARCH model with time-varying correlations is fitted to the data and three different error-distributions are considered. The performances of the portfolios are studied both in- and out-of-sample and the analysis is based on the Sharpe ratio.

The results indicate that a mean-variance portfolio based on the relationship with the macroeconomic indicators outperforms the other portfolios for the in-sample period, with respect to the Sharpe ratio. In the out-of-sample period however, none of the portfolio strategies has Sharpe ratios that are statistically different from that of an equally weighted portfolio.

Keywords: Smart Beta, Time-varying Correlations, Portfolio optimization

Sammanfattning

Den här uppsatsen undersöker möjligheten att hitta ett förhållande mellan Nasdaq Nordeas Smart Beta Index och en serie av makroekonomiska indikatorer. Detta förhållande kommer att implementeras som ett signalvärde i en portfölj bestående av alla sex index. För att se vilken påverkan signalvärdet har på en portföljs prestanda så undersöks tre portföljstrategier där den likaviktade portföljen används som riktmärke. Portföljerna ska omviktas varje månad och de portföljer som undersöks är *mean-variance* portföljen baserad på förhållandet med makroindikatorerna samt utan och den likaviktade portföljen baserad på förhållandet med makroindikatorerna.

För att kunna göra en prognos av portföljens prestanda anpassas en multivariat GARCH modell med tidsvarierande korrelationer till denna data. Tre stycken felfördelningar övervägs för modellen. Portföljens prestanda mäts både för prognosen samt för tiden med känd data och analysen är baserad på portföljernas *Sharpe ratio*.

Resultaten visar på att den portfölj som presterar bäst i tiden med känd data är *mean-variance* portföljen baserad på makroindikatorerna. I tiden för prognosen så är ingen portföljs *Sharpe ratio* statistiskt skild från den av den likaviktade portföljen.

Nyckelord: Smart Beta, Tidsvarierande korrelation, Portföljoptimering.

Acknowledgements

I would like to thank Elena Westerdahl at Nordea for supplying me with this subject and for all of her inputs and support. I would also like to thank my supervisor at KTH, Boualem Djehiche, for his help during my work.

Stockholm, September 2015

Alexandra Andersson

Contents

1	Introduction	1
1.1	Background	1
1.2	Methodology	2
2	Theoretical Background	5
2.1	Smart Beta	5
2.2	Macroeconomic Indicators	5
2.3	Test for ARCH behaviour	7
2.4	Volatility Modelling	8
2.5	Time-Varying Correlation	9
2.6	Interior-point algorithm	13
2.7	Forecasting	15
3	Data Analysis and Investment Strategy	17
3.1	Data Analysis	17
3.2	Investment Strategy	23
4	Results and Discussion	25
4.1	Model Selection	25
4.2	Forecasting the Correlation	33
4.3	Portfolio Performance	36
5	Conclusions and Remarks	39
	Appendices	44
A	Regression statistics	45

Chapter 1

Introduction

In this chapter some background to the subject is presented and the methodology of the thesis is described.

1.1 Background

In October 2014 Nordea released six Smart Beta Indices. These indices are, unlike the common practice for index weighting, not weighted on market capitalization but on the companies dividends, volatilities, momentum and the combination of these factors. The volatility index consists mostly of large capitalization stocks that has a low volatility while the momentum index consists of small capitalization stocks which may have a larger volatility but also tend to grow faster. The Dividend index consists of stocks of companies that yields large dividends. The indices have historically outperformed the market most of the time, which is why it might be beneficial to build a portfolio consisting of the outperformance of the indices against the market.

When modelling financial data the volatility plays an important role. In the classical portfolio theory introduced by Markowitz, it states that the optimal portfolio weights can be found by minimizing the variance of the portfolios returns while letting the expected portfolio return be constrained by a specified value. When implementing this strategy, the covariance matrix of the returns and the expected returns are of great importance. For financial data this covariance matrix, which depends on the correlations and the variances, is often time-dependent and the co-movements among the assets has an impact on the portfolio risks.

1.2 Methodology

The objective with this thesis is to find a relationship between the Nordea Smart Beta indices and a series of macroeconomic indicators. Out of this relationship a signal-value will be constructed. A portfolio will be constructed out of all six indices and the portfolio weights will be re-evaluated monthly based on the signal-value. An equally weighted portfolio will be used as a benchmark and a mean-variance portfolio will be used to investigate the impact of the signal-value. To determine the optimal weights for the portfolio, the relationship between the indicators and the indices must be established. The indices are assumed to follow a GARCH(p,q) process with time-varying correlations. This will be tested by the Ljung-Box test of autocorrelation, Engle's test of heteroscedasticity and by inspection of the autocorrelation function.

The GARCH process is chosen to model the volatility of the data, which is needed in order to provide a good estimate of the covariance matrix for the mean-variance portfolio. Another common way to estimate the volatility is historical simulation of the standard deviation. The historical simulation uses equal weights for all returns back to an arbitrary date while the GARCH process lets the most recent returns receive the most weight. Since financial data often contains volatility clustering and high persistence, the most recent returns should contain more information about the volatility than older returns and this is one of the advantages of the GARCH process over the historical simulation.

The first step is to identify a signal-value. This will be done graphically, where the signal-value will be estimated out of a combination of indicators. Since some of the indicators are based on a monthly survey, and the portfolio will be rebalanced at a monthly basis, monthly data of the indices will be used for the analysis. The indicators used are the Ifo Business expectations, PMI, CPI, LABOUR, GDP and the government bond with a 5 year maturity date. These are described in Section 2.2. The indicators are fitted by linear regression to the outperformance of the indices against the Stockholm Benchmark Index (SBX) to find a suitable combination. The new series', the combination of indicators, movements will be compared with the outperformance of the indices to identify a suitable signal-value for when to, or not to, invest in each index.

The portfolio consists of all six indices, where each index is assigned a weight, $\omega_i \in [0, 1]$, that will be re-evaluated monthly. It is assumed that there exists no risk-free asset. The re-evaluation of the weights will be based on two approaches; the $1/N$ rule (or naïve diversification) where the wealth is divided equally over the N assets, and the minimum variance portfolio with a lower bound for the expected portfolio returns. The two approaches will be evaluated with and without the impact of the indicators. The $1/N$ rule will be used as a benchmark to compare the other strategies with. The portfolios are used to demonstrate the impact of the signal-value, and the choices of portfolios are arbitrary. The

equally weighted portfolio is the simplest portfolio strategy, but it has also been proved to outperform all other models in out-of-sample periods [9], and the mean-variance portfolio is a well known portfolio strategy introduced by Markowitz [21].

The covariance matrix is needed to optimize the mean-variance portfolio, and this will be estimated by the GARCH(p,q) model with time-varying correlations. Since financial data often exhibit excess kurtosis, a normal distribution may not be optimal to describe the data. Three distributions will be considered for the series; the normal distribution, the Student's t -distribution, and the skew t -distribution presented by [2]. To describe the time-varying correlations, the TVC-GARCH(p,q) presented by [32] is used. The forecasting performance of each investment strategy will be analysed.

Chapter 2

Theoretical Background

2.1 Smart Beta

Alpha and beta are risk ratios used to calculate and compare returns against a benchmark, such as the S&P 500 index. Alpha is the amount that the portfolio has outperformed or underperformed the benchmark, for example if the portfolio returns were 5%, and the benchmarks returns were 2%, the alpha would be +3. Alpha strategies aim to outperform the market while beta strategies aim to follow the market. The beta represents the degree to which a portfolio is more or less volatile than the benchmark index. A beta index of 1 implies that the fund will move with the market, smaller than 1 indicates a fund less volatile and larger than 1 is a fund more volatile than the market [27].

Smart beta strategies (also known as e.g "alternative beta" or "advanced beta") are strategies that tracks indices where the weights of the assets are not assigned based on the size or price of the instrument. This kind of weighting scheme was introduced as the fundamental indexation by Robert Arnott et al. in 2005, where the assets where weighted by book value, revenues, dividends and others. The results were successful with a higher return and lower variance than capitalization weighted indices [1].

The Nasdaq Nordea Smart Beta Indices are based on three different weight components and the combination of these; volatility, dividends, momentum, volatility-momentum, volatility-dividend and dividend-momentum. Each of these six indices contains 30 stocks that are listed at Nasdaq Stockholm, and they are selected and weighted at a quarterly basis [25].

2.2 Macroeconomic Indicators

Macroeconomic indicators are statistics that are used to indicate the current status in the economy, or how the economy will develop in the near future in a specific area. In the succeeding sections six macroeconomic indicators are explained. The government bond

with a 5 year maturity date will also be considered as an indicator.

2.2.1 Ifo Business Climate

The Ifo Business Climate is based on approximately 7,000 monthly survey responses of firms in the market of manufacturing, construction, wholesaling and retailing in Germany. The surveys investigate the firms current business situations and their expectations for the following six months. There are three different responses to evaluate their current situation; *good*, *satisfactory* or *poor* and their expectations are categorized as *more favourable*, *unchanged* or *more unfavourable*. The replies are weighted according to the importance of the industry and the Business Climate is the mean of the balances of the current situation and the expectations. The balance value of the current situation is the difference in the percentages of the responses *good* and *poor* and the balance value for the expectations is the average of the responses *more favourable* and *more unfavourable*. The Ifo Business Climate fluctuates between values of -100 (corresponding to all companies evaluating their situation as poor or expecting the business to become worse) and 100 (corresponding to all companies evaluating their situation as good or expecting an improvement) [15].

2.2.2 PMI

The PMI is an indicator of the manufacturing sector and it is based on five major indicators; inventory levels, production, supplier deliveries and the employment environment. The index derives from a monthly survey, where data is collected from 400 purchasing managers, where the respondents state whether business conditions has *improved*, *deteriorated* or *stayed the same* compared with the previous month [20].

The PMI has a base value of 50 and a value higher than that indicates an expansion of the manufacturing sector compared to the previous month. Values lower than 50 represents a contraction while a value of 50 indicates no change [30].

2.2.3 GDP

The gross domestic product (GDP) is the value of all goods and services that are produced in the country with the purpose of being used for consumption, export and investment, normally during a period of a year or quarter. It can be calculated in three ways [29]:

- As the sum of the final use of the goods and services, while adding the export and subtracting the import of goods and services.
- As the sum of gross value of the various institutional industries, adding the taxes and subtracting the subsidies.
- As the sum of usage in the account of the total economies generation of income.

2.2.4 CPI

The consumer price index (CPI) measures the prices that the consumers actually pay, i.e. including indirect taxes and subventions. In Sweden, the prices are gathered during three weeks every month and are weighted to see how big of an impact each product has on the domestic consumption. The index measure the change in the prices since the previous month with a base value of 100 [28].

2.2.5 Labour Participation Rate

The labour participation rate is the ratio between the labour force and the overall size of their cohort. The labour force in Sweden is defined as the number of the population between the ages of 15-74 who are employed, and those who are unemployed and searching for a job [14].

2.3 Test for ARCH behaviour

To test the data for ARCH behaviour, the sample autocorrelation is investigated. A series exhibiting ARCH behaviour typically has zero autocorrelation while the autocorrelations of the squared returns are significantly different from zero. The Ljung-Box test and Engle's test for residual heteroscedasticity are applied to confirm the results from the analysis of the sample autocorrelation.

2.3.1 The Ljung-Box Q-test

The Ljung-Box Q-test uses a single test statistic:

$$Q = n(n+2) \sum_{j=1}^h \frac{\hat{\rho}^2(j)}{n-j} \quad (2.1)$$

where $\hat{\rho}^2(j)$ is the squared autocorrelations of the process X_t at lag j . If the process X_1, \dots, X_m is a finite-variance IID sequence, then Q is approximately distributed as the sum of squares of the independent $N(0, 1)$ variables, i.e. chi-squared with m degrees of freedom. A large value of Q suggest that the sample autocorrelations of the data are too large to come from an IID sequence. Therefore the hypothesis is rejected if $Q > \chi_{1-\alpha}^2(m)$, where $\chi_{1-\alpha}^2(m)$ is the $1 - \alpha$ quantile of the chi-squared distribution with m degrees of freedom [19].

2.3.2 Engle's test for residual heteroscedasticity

Engle's test for residual heteroscedasticity is a Lagrange multiplier test that tests the null hypothesis that the squared residuals are white noise against the alternative hypothesis

that there is autocorrelation in the squared residuals. This is equivalent to the test statistic F for testing $\alpha_i = 0$ in the linear regression [10]:

$$a_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \cdots + \alpha_m a_{t-m}^2 + e_t, \quad t = m + 1, \dots, T$$

where m is a pre-specified positive integer, T is the sample size and e_t denotes the error term. The null hypothesis is $H_0 : \alpha_0 = \cdots = \alpha_m = 0$. The test statistic in Engle's ARCH test is the usual F statistic for the regression on the squared residuals. Under the null hypothesis the F statistic follows a χ^2 distribution with m degrees of freedom. The hypothesis is rejected if $F > \chi_{1-\alpha}^2(m)$, where $\chi_{1-\alpha}^2(m)$ is the $1 - \alpha$ quantile of the χ^2 distribution with m degrees of freedom

2.4 Volatility Modelling

Financial time series tend to have large fluctuations, a period of calm behaviour tend to be followed by calm behaviour while large changes tend to follow large changes. Therefore it is important to model the data with respect to the volatility in order to account for these large fluctuations. Let r_t be the return of the data y_t , defined as:

$$r_t = 100 \log \left(\frac{y_t}{y_{t-1}} \right). \quad (2.2)$$

Then the conditional variance and the conditional expected return is defined as:

$$\mu_t = E[r_t | F_{t-1}], \quad \sigma_t^2 = Var[r_t | F_{t-1}] = E[(r_t - \mu_t)^2 | F_{t-1}],$$

where F_{t-1} is the information available at time $t - 1$. To reduce the number of parameters in need to be estimated, the conditional expected returns are replaced with the sample mean, defined as:

$$\bar{\mu} = \frac{1}{T} \sum_{t=1}^T r_t \quad (2.3)$$

2.4.1 Univariate GARCH(p,q)

The Auto-Regressive Conditional Heteroscedasticity (ARCH) process models the volatility clustering where the conditional variance depends on lagged values of the squared returns [10]. Let

$$a_t = r_t - \mu_t, \quad (2.4)$$

define the unexpected returns. Then the process $\{a_t, t \in \mathbb{Z}\}$ is said to be an ARCH(p) process if it can be written as:

$$a_t = \sigma_t z_t, \quad z_t \sim IID(0, 1) \quad (2.5)$$

where

$$\sigma^2 = \alpha_0 + \sum_{i=1}^p \beta_i a_{t-1}^2 \quad (2.6)$$

and $\alpha_0 > 0, \beta_i \geq 0$ and if z_t and a_t are independent for all t . However, the ARCH(p) process tend to require a large number of parameters to correctly describe the process [31]. In 1986, Bollerslev introduced a generalized autoregressive model (GARCH) [4] that requires far less parameters to correctly describe a volatility process. It also models the conditional variance with lagged values of the conditional variance as well as lagged values of the squared returns. Let a_t be defined as in (2.4), then a_t is said to be a GARCH(p,q) process if it can be written as:

$$a_t = \sigma_t z_t, \quad z_t \sim IID(0, 1) \quad (2.7)$$

where

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \beta_i a_{t-1}^2 + \sum_{j=1}^q \alpha_j \sigma_{t-j}^2 \quad (2.8)$$

and $\alpha_0, \alpha_i, \beta_j \geq 0$ and $\sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_j) < 1$ and z_t and a_t are independent for all t . The second constraint implies that the unconditional variance of a_t^2 is finite whereas the conditional variance of σ_t^2 evolves over time.

2.4.2 Multivariate GARCH(p,q)

Consider N assets with their t -days returns in a $N \times 1$ vector \mathbf{r}_t . Then the returns can be represented by:

$$\mathbf{r}_t = \boldsymbol{\mu} + \mathbf{a}_t \quad (2.9)$$

where

$$\mathbf{a}_t = \mathbf{H}^{1/2} \mathbf{z}_t. \quad (2.10)$$

Here \mathbf{H}_t is the conditional covariance matrix of \mathbf{a}_t and $\mathbf{H}_t^{1/2}$ its Cholesky factorization, $\boldsymbol{\mu}_t$ is the vector of expected returns, \mathbf{a}_t the vector of unexpected returns and \mathbf{z}_t is the vector of innovations with $E[\mathbf{z}_t] = \mathbf{0}$ and $Var[\mathbf{z}_t] = \mathbf{I}_N$, the identity matrix.

In this multivariate framework, each conditional variance $\sigma_{i,t}^2$ is assumed to follow a univariate GARCH(p,q) process as defined in 2.4.1. It can be noted that the order (p,q) may be varied with i for each conditional variance term $\sigma_{i,t}^2$ [32]. The modelling of the conditional variance matrix is presented in Section 2.5.

2.5 Time-Varying Correlation

Let N be the number of assets and $\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t$, where \mathbf{D}_t is the diagonal matrix with $diag(\mathbf{D}_t) = (\sigma_{1,t}, \dots, \sigma_{N,t})$ and \mathbf{R}_t is the correlation matrix of \mathbf{a}_t . Then ε_t is defined as the

GARCH residuals:

$$\varepsilon_t = \mathbf{D}_t^{-1} \mathbf{a}_t. \quad (2.11)$$

There are several methods to model the conditional correlation, see for example [11], [5] and [32]. The constant conditional correlation (CCC) model by Bollerslev [5], assumes that all correlations are time-invariant. In 2002, Engle proposed the dynamic conditional correlation (DCC) model [11], which allows the correlations to be time-varying. In this thesis the correlations are assumed to be time-varying and the time-varying correlation (TVC) model of Tse and Tsui (2002) [32] will be used to describe the conditional covariance matrix. In the TVC framework, the conditional correlations are assumed to follow an ARMA process defined as:

$$R_t = (1 - \theta_1 - \theta_2) \bar{R} + \theta_1 R_{t-1} + \theta_2 \Psi_{t-1}. \quad (2.12)$$

Here \bar{R} is a time invariant correlation matrix. Ψ_t is the sample correlation of $\{\varepsilon_{t-1}, \dots, \varepsilon_{t-M}\}$ defined as:

$$\psi_{ij,t-1} = \frac{\sum_{h=1}^M \varepsilon_{i,t-h} \varepsilon_{j,t-h}}{\sqrt{\left(\sum_{h=1}^M \varepsilon_{i,t-h}^2\right) \left(\sum_{h=1}^M \varepsilon_{j,t-h}^2\right)}}, \quad 1 \leq i \leq j \leq N. \quad (2.13)$$

Thus the correlation will be a weighted average of \bar{R}_t , R_{t-1} and Ψ_{t-1} which differs it from the correlation defined in the DCC model which only depends on the single lagged term $\varepsilon_{t-1} \varepsilon_{jt-1}$, see [11].

Define $E_t = (\varepsilon_{t-1}, \dots, \varepsilon_{t-M})$ and $B_t = \left(\sum_{h=1}^M \varepsilon_{i,t-h}^2\right)^{1/2}$ for $i = 1, \dots, N$, then:

$$\Psi_{t-1} = B_{t-1}^{-1} E_{t-1} E'_{t-1} B_{t-1}^{-1}. \quad (2.14)$$

The constraint $M \geq N$ ensures that the matrix Ψ_t is positive-definite and Ψ_t is a well defined correlation matrix. In this thesis the constraint $M = N$ will be used, following the example of Tse and Tsui.

2.5.1 The distribution of the innovations & parameter estimation

The innovations z_t are assumed to be IID with $E[\mathbf{z}_t] = \mathbf{0}$ and $Var[\mathbf{z}_t] = \mathbf{I}_N$, the distribution however remains to be determined. In this thesis, three different distributions will be considered: the Gaussian distribution, the Student's t -distribution and the skew Student's t -distribution. The Student's t -distribution is chosen since financial data is known to have excess kurtosis and the Student's t -distribution is able to capture heavy tails, unlike the Gaussian distribution. Though the Student's t -distribution can capture heavy tails it is restricted to symmetric distribution around the mean. Therefore, the skew Student's t -distribution introduced in [2] is used to model the (possible) skewness and kurtosis of the

returns.

The parameters, θ , to be estimated in the TVC model will depend on the distribution for the innovation and will be estimated by maximum-likelihood estimation. The likelihood function is defined as

$$L(a_1, \dots, a_t | \theta) = \prod_{t=1}^T \varphi(a_t), \quad (2.15)$$

where $\varphi(a_t)$ is the density function of the unexpected returns a_t .

Multivariate Gaussian distributed innovations

When the innovations, \mathbf{z}_t , are assumed to be Gaussian distributed, their joint density function is defined as:

$$\varphi(\mathbf{z}_t) = \prod_{t=1}^T \frac{1}{(2\pi)^{N/2} |\mathbf{H}_t|^{1/2}} e^{-\frac{1}{2} \mathbf{z}_t' \mathbf{z}_t}. \quad (2.16)$$

Then the likelihood function of $\mathbf{a}_t = \mathbf{H}_t^{1/2} \mathbf{z}_t$ is:

$$L(\mathbf{a}_t | \theta) = \prod_{t=1}^T \frac{1}{(2\pi)^{N/2} |\mathbf{H}_t|^{1/2}} e^{-\frac{1}{2} \mathbf{a}_t' \mathbf{H}_t^{-1} \mathbf{a}_t}, \quad (2.17)$$

where $\theta_i = (\alpha_{i,0}, \alpha_{i,1}, \dots, \alpha_{i,p}, \beta_{i,1}, \dots, \beta_{i,q}, \theta_1, \theta_2)$ are the parameters to be estimated for $i = 1, \dots, N$. Maximizing the likelihood function yields the same result as maximizing the logarithm of the likelihood function, doing this and substituting $\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t$ yields

$$\begin{aligned} \log L(\mathbf{a}_t | \theta) &= -\frac{1}{2} \sum_{t=1}^T (N \log(2\pi) + \log(|\mathbf{H}_t|) + \mathbf{a}_t' \mathbf{H}_t^{-1} \mathbf{a}_t) \\ &= -\frac{1}{2} \sum_{t=1}^T (N \log(2\pi) + \log(|\mathbf{D}_t \mathbf{R}_t \mathbf{D}_t|) + \mathbf{a}_t' \mathbf{D}_t^{-1} \mathbf{R}_t^{-1} \mathbf{D}_t^{-1} \mathbf{a}_t) \end{aligned} \quad (2.18)$$

Multivariate Student's t -distributed innovations

When the innovations are assumed to follow a Student's t -distribution, the joint density function of z_1, \dots, z_t with ν degrees of freedom is defined as:

$$\varphi(\mathbf{z}_t) = \prod_{t=1}^T \frac{\Gamma(\frac{\nu+N}{2})}{\Gamma(\frac{\nu}{2}) |\mathbf{H}_t|^{1/2} [\pi(\nu-2)]^{N/2}} \left[1 + \frac{\mathbf{z}_t' \mathbf{z}_t}{\nu-2} \right]^{-\frac{N+\nu}{2}}, \quad (2.19)$$

where $\Gamma(\cdot)$ is the Gamma function. Then the likelihood function of $\mathbf{a}_t = \mathbf{H}_t^{1/2} \mathbf{z}_t$ is defined as:

$$L(\mathbf{a}_t | \theta) = \prod_{t=1}^T \frac{\Gamma\left(\frac{\nu+2}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) [\pi(\nu-2)]^{N/2} |\mathbf{H}_t|^{1/2}} \left[1 + \frac{\mathbf{a}_t^T \mathbf{H}_t^{-1} \mathbf{a}_t}{\nu-2}\right]^{-\frac{N+\nu}{2}}, \quad (2.20)$$

where $\theta_i = (\alpha_{i,0}, \alpha_{i,1}, \dots, \alpha_{i,p}, \beta_{i,1}, \dots, \beta_{i,q}, \theta_1, \theta_2, \nu)$ are the parameters to be estimated. When taking the logarithm and substituting $\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t$ the following log-likelihood function of observation t is obtained:

$$\begin{aligned} \log L_t(\mathbf{a}_t | \theta) &= \log \left[\Gamma\left(\frac{\nu+N}{2}\right) \right] - \log \left[\Gamma\left(\frac{\nu}{2}\right) \right] - \frac{N}{2} \log[\pi(\nu-2)] \\ &\quad - \frac{1}{2} \log \left[|\mathbf{D}_t \mathbf{R}_t \mathbf{D}_t| \right] - \frac{\nu+N}{2} \log \left[1 + \frac{\mathbf{a}_t^T \mathbf{D}_t^{-1} \mathbf{R}_t^{-1} \mathbf{D}_t^{-1} \mathbf{a}_t}{\nu-2} \right]. \end{aligned} \quad (2.21)$$

Then the log-likelihood function of the whole sample can be composed by the sum of each log-likelihood function at time t :

$$\log L(\mathbf{a}_t | \theta) = \sum_{t=1}^T \log L_t. \quad (2.22)$$

Multivariate Skew Student's t -distributed innovations

The multivariate skew Student's t -distribution was introduced in [2] since financial data often have a coefficient of kurtosis larger than 3 (the case for the normal distribution) and are often skewed. The joint density of the skew t -distribution, with the coefficients of skewness $\xi = (\xi_1, \dots, \xi_N)$ and ν degrees of freedom is:

$$\varphi(\mathbf{z}_t) = \left(\frac{2}{\sqrt{p_i}}\right)^N \left(\prod_{i=1}^N \frac{\xi_i s_i}{1 + \xi_i^2}\right) \frac{\Gamma\left(\frac{\nu+2}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) |\mathbf{H}_t|^{1/2} (\nu-2)^{N/2}} \prod_{t=1}^T \left(1 + \frac{\mathbf{z}_t^{*T} \mathbf{z}_t^*}{\nu-2}\right)^{-\frac{N+\nu}{2}}, \quad (2.23)$$

where

$$z_t^* = (s_i z_t + m_i) \xi_i^{I_i}, \quad (2.24)$$

$$m_i = \frac{\Gamma\left(\frac{\nu-1}{2}\right) \sqrt{\nu-2}}{\sqrt{\pi} \Gamma\left(\frac{\nu}{2}\right)} \left(\xi_i - \frac{1}{\xi_i}\right), \quad (2.25)$$

$$s_i^2 = \left(\xi_i^2 + \frac{1}{\xi_i^2} - 1\right) - m_i^2 \quad (2.26)$$

and

$$I_i = \begin{cases} 1 & \text{if } z_i < \frac{m_i}{s_i} \\ -1 & \text{if } z_i \geq \frac{m_i}{s_i}. \end{cases} \quad (2.27)$$

Note that the variables s_i and m_i are functions of ξ and ν and does not represent additional parameters. The density function is well defined for $\xi > 0$ and $\nu > 2$. The model parameters to be estimated when assuming the innovations are skew t -distributed are $\theta_i = (\alpha_{i,0}, \alpha_{i,1}, \dots, \alpha_{i,p}, \beta_{i,1}, \dots, \beta_{i,q}, \theta_1, \theta_2, \nu, \xi_i)$. A value of $\log(\xi - i) < 0$ represents a distribution skewed to the left and if $\log(\xi - i) > 0$, the distribution is right-skewed. If $\log(\xi - i) = 0$ the distribution is symmetric.

The likelihood function of $\mathbf{a}_t = \mathbf{H}_t^{1/2} \mathbf{z}_t$ is defined as:

$$L(\mathbf{a}_t | \theta) = \prod_{t=1}^T \frac{\Gamma(\frac{\nu+N}{2})}{|\mathbf{H}|^{1/2} \Gamma(\frac{\nu}{2}) (\nu-2)^{N/2}} \prod_{i=1}^N \frac{\xi_i s_i}{1 + \xi_i^2} \left[1 + \frac{(\mathbf{H}^{-1/2} \mathbf{a}_t)^* T (\mathbf{H}^{-1/2} \mathbf{a}_t)^*}{\nu-2} \right]^{-\frac{\nu+N}{2}} \quad (2.28)$$

2.5.2 Maximizing the likelihood-function

The maximization of the likelihood function can, if assuming a normal distribution of the innovations, be determined by a two-step approach as suggested by Engle (2002) [11]. This approach consists of splitting the likelihood function in two parts, the first for a series of univariate GARCH estimation and the second the correlation estimation. This two-step approach can be approximated for the t -distribution and skewed t -distribution by assuming normal distributed innovations when calculating the volatility parameters and then using a quasi maximum-likelihood estimation for the correlation coefficients. This has proven to give similar estimates of the parameters as when estimating both the volatility and the correlation parameters in one step [2]. In this thesis, the one-step approach will be used for all distributions, including the Gaussian assumption. The optimization will be conducted by using the matlab function `fmincon()` [22], this algorithm is described in Section 2.6.

2.6 Interior-point algorithm

The interior-point algorithm approximates the original problem

$$\begin{aligned} & \underset{\mathbf{x}}{\text{minimize}} && f(\mathbf{x}) \\ & \text{subject to} && h(\mathbf{x}) = \mathbf{0}, \\ & && g(\mathbf{x}) \leq \mathbf{0}, \end{aligned} \quad (2.29)$$

by introducing slack variables $s = (s_1, \dots, s_n)$, $s_i \geq 0$. The number of slack variables equals the number of constraints g and the approximated problem is formulated as:

$$\begin{aligned} & \underset{\mathbf{x}, \mathbf{s}}{\text{minimize}} && f(\mathbf{x}) - \mu \sum_{i=1}^n \ln(s_i) \\ & \text{subject to} && h(\mathbf{x}) = \mathbf{0}, \\ & && g(\mathbf{x}) + s = \mathbf{0}. \end{aligned} \quad (2.30)$$

As μ decreases to zero, the minimum of f_μ should approach the minimum of f . To solve the approximate problem the algorithm uses either a *direct step* or a *conjugate gradient (CG) step*.

The direct step attempts to solve the KKT equations via linear approximations. The KKT equations are first order necessary conditions for a non-linear programming solution to be optimal and they are defined as:

$$\nabla f(\mathbf{x}) + \sum_{k=1}^m \lambda_k \nabla g_k(\mathbf{x}) + \sum_{j=1}^m y_j \nabla h(\mathbf{x}) = \mathbf{0}, \quad (2.31)$$

$$\lambda_k g_k(\mathbf{x}) = 0, \quad k = 1, \dots, m, \quad (2.32)$$

$$g_k(\mathbf{x}) \leq 0, \quad k = 1, \dots, m, \quad (2.33)$$

$$h_k(\mathbf{x}) = 0, \quad k = 1, \dots, m, \quad (2.34)$$

$$\lambda_k \geq 0, \quad k = 1, \dots, m. \quad (2.35)$$

The direct step $(\Delta x, \Delta s)$ is defined as:

$$\begin{bmatrix} H & 0 & J_h^T & J_g^T \\ 0 & S\Lambda & 0 & -S \\ J_h & 0 & I & 0 \\ J_g & -S & 0 & I \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta s \\ -\Delta y \\ -\Delta \lambda \end{bmatrix} = - \begin{bmatrix} \nabla f - J_h^T y - J_g^T \lambda \\ S\lambda - \mu e \\ h \\ g + s \end{bmatrix}.$$

where

- J_g is the Jacobian of the constraint function g ,
- J_h is the constraint function h , $S = \text{diag}(s)$,
- λ denotes the Lagrange multiplier vector associated with constraint g ,
- $\Lambda = \text{diag}(\lambda)$,
- y denotes the Lagrange multiplier vector associated with constraint h
- and e is a vector of ones.

The algorithm starts by attempting to take a direct step to minimize the merit function $f_\mu(x, s) + \nu \| (h(x), g(x) + s) \|^2$. If the step does not decrease the merit function, a new step is attempted. If the algorithm can not take a direct step, for example if the approximate problem is not locally convex near the current iterate, the algorithm takes a CG step. The CG step minimizes a quadratic approximation to the approximate problem in a trust

region, subject to linearized constraints. Let R define the radius of the trust region, then the algorithm solves the following KKT equations in the least-squares sense:

$$\nabla_x L = \nabla_x f(\mathbf{x}) + \sum_k \lambda_k \nabla g_k(\mathbf{x}) + \sum_j y_j \nabla h_j(\mathbf{x}), \quad (2.36)$$

subject to λ being positive. The algorithm then takes the step $(\Delta \mathbf{x}, \Delta \mathbf{s})$ to approximately solve the problem:

$$\begin{aligned} & \underset{\Delta \mathbf{x}, \Delta \mathbf{s}}{\text{minimize}} && \nabla f^T \Delta \mathbf{x} + \frac{1}{2} \Delta \mathbf{x}^T \nabla_{xx}^2 L \Delta \mathbf{x} + \mu e^T S^{-1} \mathbf{s} + \frac{1}{2} \mathbf{s}^T S^{-1} \Lambda \mathbf{s} \\ & \text{subject to} && g(\mathbf{x}) + J_g \Delta \mathbf{x} + \Delta \mathbf{s} = \mathbf{0}, \\ & && h(\mathbf{x}) + J_h \Delta \mathbf{x} = 0. \end{aligned} \quad (2.37)$$

The algorithm attempts to minimize the norm of the linearized constraint, and will stop when either the maximum number of iterations is reached or when the first-order optimality measure is within its tolerance level [23].

2.7 Forecasting

To minimize the mean-variance portfolio, the forecast of the covariance matrix \mathbf{H}_{t+1} is needed. This forecast can be conducted in two steps since $\mathbf{H}_{t+1} = \mathbf{D}_{t+1} \mathbf{R}_{t+1} \mathbf{D}_{t+1}$. First \mathbf{D}_{t+1} can be estimated by the forecast of σ_{t+1}^2 , then the correlation matrix \mathbf{R}_{t+1} can be estimated with the forecast of the conditional variance and the estimate ρ_{t+1} .

2.7.1 Forecasting the conditional variance

The unconditional variance of a GARCH(1,1) process is defined as:

$$\begin{aligned} \text{Var}(a_t) &= E[a_t^2] - E[a_t]^2 \\ &= E[a_t^2] \\ &= E[\sigma_t^2 z_t^2] \\ &= \{\text{Since } \sigma_t \text{ and } z_t \text{ are independent}\} \\ &= E[\sigma_t^2] E[z_t^2] \\ &= E[\sigma_t^2] \\ &= \alpha_0 + \alpha_1 E[a_{t-1}^2] + \beta_1 E[\sigma_{t-1}^2] \\ &= \alpha_0 + (\alpha_1 + \beta_1) \text{Var}[a_{t-1}] \end{aligned} \quad (2.38)$$

and since a_t is a stationary process,

$$\text{Var}(a_t) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}. \quad (2.39)$$

The k -step ahead forecast of the conditional variances, $\hat{\sigma}_{t+k}^2$ for a GARCH(1,1) process can be derived from the definition of σ_t^2 :

$$\begin{aligned}
E[\hat{\sigma}_{t+k}^2|F_{t+k-1}] &= \alpha_0 + \alpha_1 E[a_{t+k-1}^2|F_{t+k-1}] + \beta_1 E[\sigma_{t+k-1}^2|F_{t+k-1}] \\
&= \alpha_0 + \alpha_1 E[a_{t+k-1}^2] + \beta_1 E[\sigma_{t+k-1}^2] \\
&= \alpha_0 + (\alpha_1 + \beta_1) E[\sigma_{t+k-1}^2] \\
&= \alpha_0 + (\alpha_1 + \beta_1) E[\alpha_0 + (\alpha_1 + \beta_1) E[\sigma_{t+k-2}^2|F_{t+k-2}]] \\
&= \dots \\
&= \sum_{i=0}^{k-2} \alpha_0 (\alpha_1 + \beta_1)^i + (\alpha_1 + \beta_1)^{k-1} E[\sigma_{t+1}^2|F_t]
\end{aligned} \tag{2.40}$$

where

$$E[\sigma_{t+1}^2|F_t] = \alpha_0 + \alpha_1 a_t^2 + \beta_1 \sigma_t^2. \tag{2.41}$$

As k increases to infinity, the estimated k step ahead forecast $\hat{\sigma}_{t+k}$ will converge to the unconditional variance (2.39).

2.7.2 Forecasting the conditional correlation

The conditional correlation as defined by Tse and Tsui is a non-linear process, thus it cannot be directly solved forward to provide a method for forecasting. By following the example of Engle and Sheppard (2001) [12], the term $E[\Psi_t]$ will be approximated by $E[\Psi_t] \approx E[R_t]$. Then the k -step ahead forecast of the conditional correlation is approximated as:

$$\begin{aligned}
E[R_{t+k}|F_{t+k-1}] &= (1 - \theta_1 - \theta_2) \bar{R} + \theta_1 E[R_{t+k-1}|F_{t+k-1}] + \theta_2 E[\Psi_{t+k-1}|F_{t+k-1}] \\
&\approx (1 - \theta_1 - \theta_2) \bar{R} + (\theta_1 + \theta_2) E[R_{t+k-1}] \\
&= (1 - \theta_1 - \theta_2) \bar{R} + (\theta_1 + \theta_2) E[(1 - \theta_1 - \theta_2) \bar{R} + (\theta_1 + \theta_2) E[R_{t+k-2}|F_{t+k-2}]] \\
&= \dots \\
&= \sum_{i=0}^{k-2} (1 - \theta_1 - \theta_2) \bar{R} (\theta_1 + \theta_2)^i + (\theta_1 + \theta_2)^{k-1} E[R_{t+1}|F_t]
\end{aligned} \tag{2.42}$$

where $E[R_{t+1}|F_t] = (1 - \theta_1 - \theta_2) \bar{R} + \theta_1 R_t + \theta_2 \Psi_t$. It can be seen that the k step ahead forecast decays with a ratio $(\theta_1 + \theta_2)$ and as k increases to infinity, the estimate \hat{R}_{t+k} will converge to the unconditional correlation matrix \bar{R} .

Chapter 3

Data Analysis and Investment Strategy

In this chapter, first the analysis of the data is described; how the signal-value is determined and how the GARCH(p,q) model is estimated. Then the investment strategy used for the portfolio is presented.

3.1 Data Analysis

In this section the analysis of the monthly data of indices and indicators as well as the daily data for the GARCH(p,q) model is presented. The data used is daily data of the Nordea Smart Beta Indices from 2004-06-30 to 2014-11-30.

3.1.1 Signal-Value

To find a signal-value for the indices, the performance of the indicators is compared to the outperformance of the indices against SBX. As can be seen in Figure 3.1, the outperformance is, except for in the beginning for the momentum index, positive. This means that so far, the indices have performed above the market which could make it beneficial to compose a portfolio consisting of the indices.

When inspecting the outperformances, it can be seen that all indices show an increasing trend. This trend is removed by approximating it with a 3rd degree polynomial. The detrended data and the trend can be seen in Figure 3.2. From the detrended data, it can also be seen that the volatility and the momentum-volatility seem to perform the best in a crisis, which is to be expected since the volatility index is designed to perform well when the market is volatile. These are also the indices with the lowest maximum drawdown, see Table 3.1.

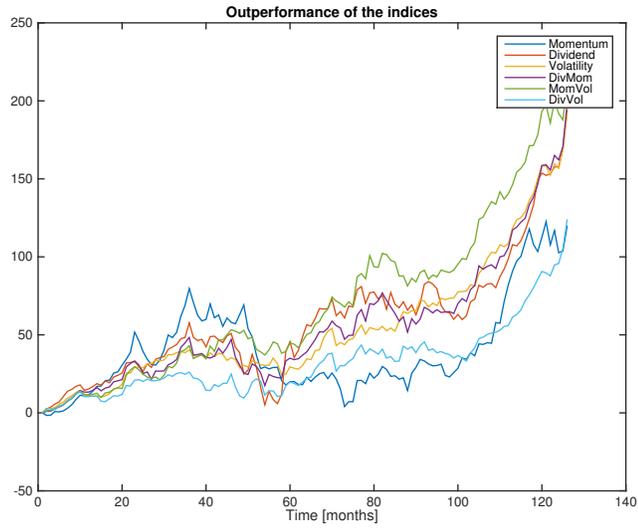


Figure 3.1: *The outperformance of the monthly data for the indices against SBX.*

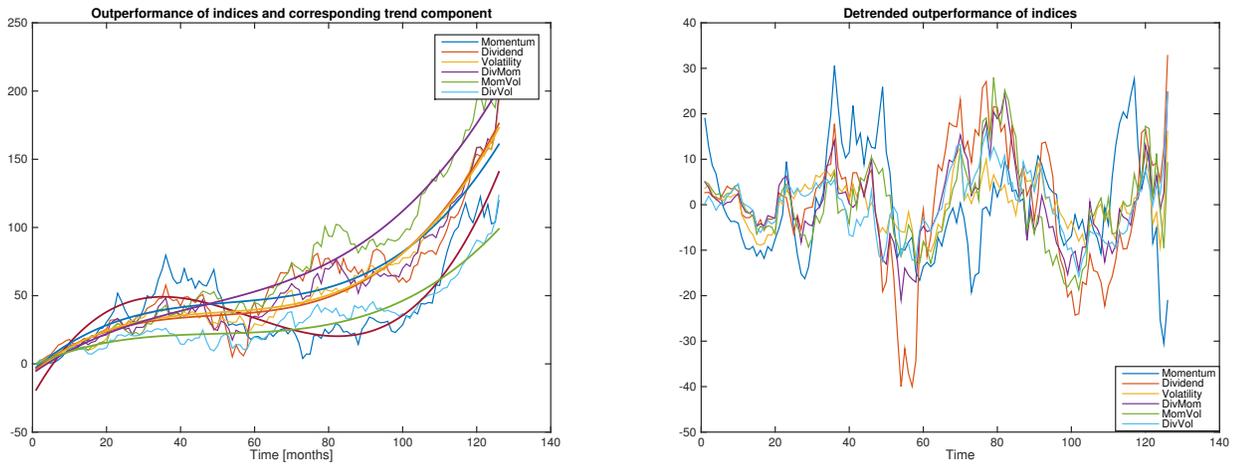


Figure 3.2: *Left: The outperformance of each index with corresponding trend component. Right: The detrended outperformance.*

Outperformance	Maximum Drawdown (%)	Index
Momentum Index	0.9492	33-70
Dividend Index	0.9102	33-51
Volatility Index	0.3926	33-55
Dividend-Momentum Index	0.6379	33-51
Momentum-Volatility Index	0.3044	43-51
Dividend-Volatility Index	0.6358	33-46

Table 3.1: *Maximum Drawdown with corresponding indices for the monthly outperformance (not returns).*

To find a suitable combination of indicators to explain the behaviour of each index, a linear regression is performed on the indices with all of the indicators to see which indicator might be statistically significant. The indicators that are not significant at a 10 % significance-level will be removed and the regression will be performed once again without the insignificant indicators. The result and statistics from the first and second regression for all indices can be seen in Appendix A. In Figure 3.3, the linear regression and the corresponding index can be seen. From the graphs it can be seen that the regression seems to be a good fit for the momentum, dividend, dividend-momentum and dividend-volatility indices but not for the volatility or momentum-volatility indices.

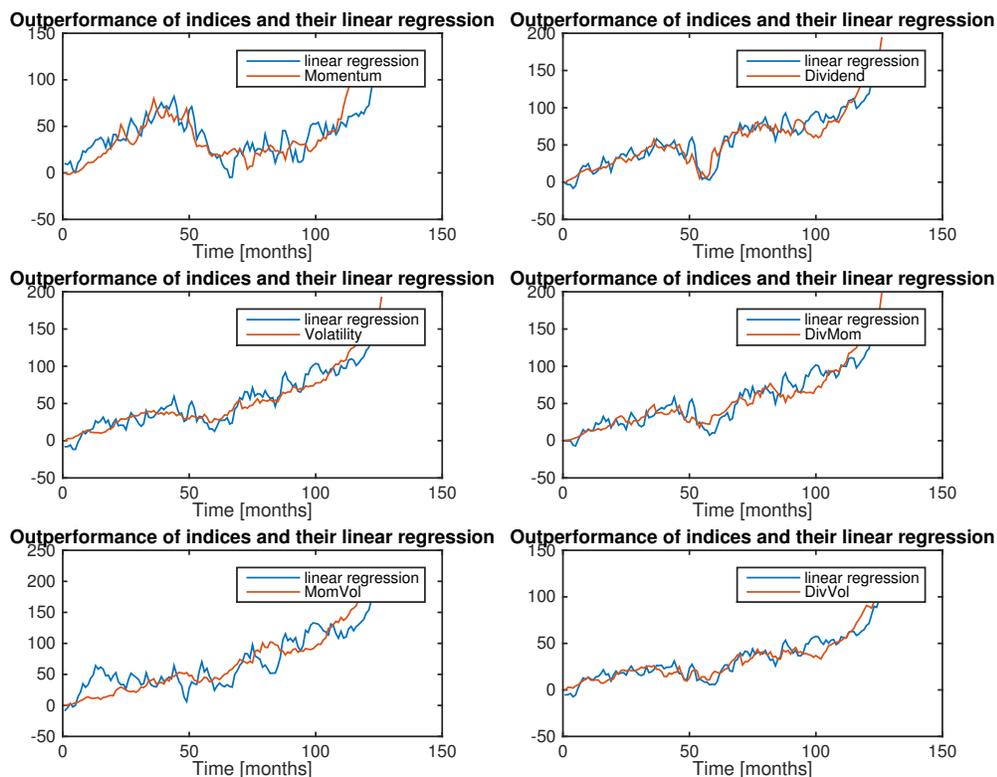


Figure 3.3: *The outperformance of the indices and their corresponding linear regression of indicators. **From the upper left:** Momentum, Dividend, Volatility, Dividend-Momentum, Momentum-Volatility, Dividend-Volatility.*

To get a clearer view of how the the indices vary with the linear regression of the indicators, the detrended data will be analysed. When looking at the graphs in Figure 3.4, it can be seen that the combination of indicators for the dividend index seem to be the best fit to the data. For the momentum, dividend, dividend-momentum and dividend-volatility index the indicator seems to follow the peaks and troughs for the indices.

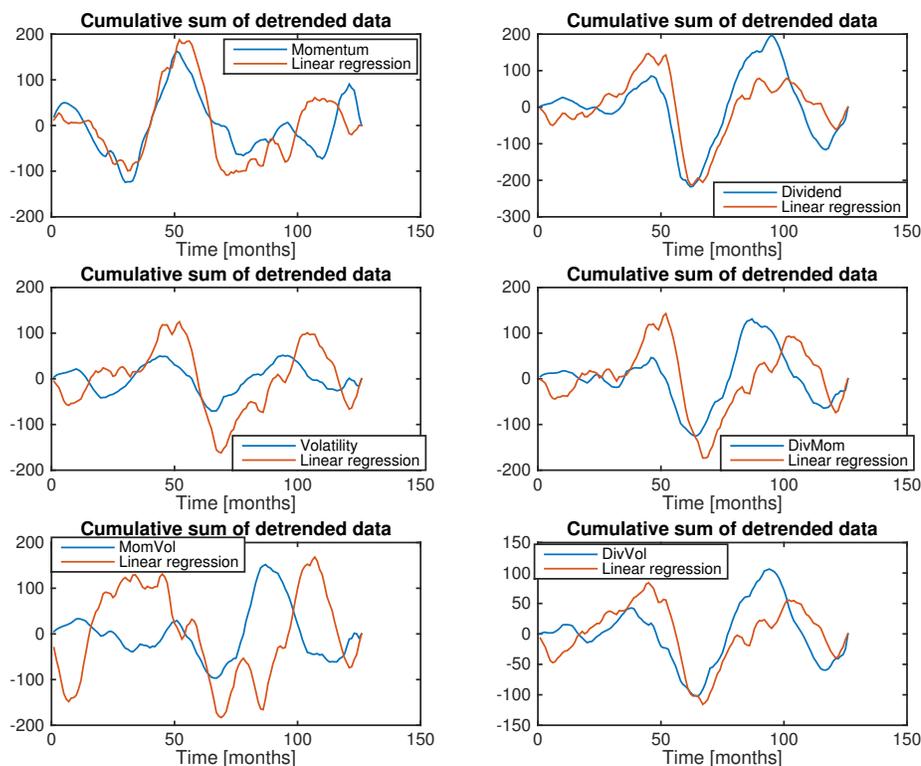


Figure 3.4: *The cumulative sum of the detrended data for the outperformance and the linear combination of indicators. **From the upper left:** Momentum, Dividend, Volatility, Dividend-Momentum, Momentum-Volatility, Dividend-Volatility.*

By inspection of the outperformance of the indices, and the linear combination of the macroeconomic indicators (which seem to indicate the large peaks and troughs), it could be possible to avoid the deepest troughs. This might be achieved by only investing in the momentum-volatility index when the linear regression indicates that the market is shifting downwards. Momentum-volatility is also the index with the fastest recovery from a downward shift, if looking at the index of the maximum drawdown in Table 3.1.

The momentum and the dividend indices are the ones that performs worst when the

market drops, and according to their respective linear regression of indicators, they drop simultaneously. The linear regression for the dividend index has the lowest root mean squared error of these two (see Appendix A), therefore the combination of indicators regressed on the dividend index will be used as an indicator for when the indices are about to drop. When the linear combination has moved one standard deviation from a peak, all the wealth will be invested in the momentum-volatility index. When the linear combination has increased two standard deviations from the trough, the mean-variance or naïve diversification strategy will be implemented again.

3.1.2 Sample autocorrelation

To check if the data is stationary, the autocorrelation function is calculated for the residuals of the outperformance of each index. As can be seen in Figure 3.5, the autocorrelations are not significantly different from zero for either of the indices, indicating that the data is independently distributed. However, when looking at the autocorrelation functions for the squared residuals (see Figure 3.6) they are significantly different from zero, indicating dependence in the data. These properties indicates that there is heteroscedasticity in the time series, therefore a GARCH(p,q) model might be appropriate.

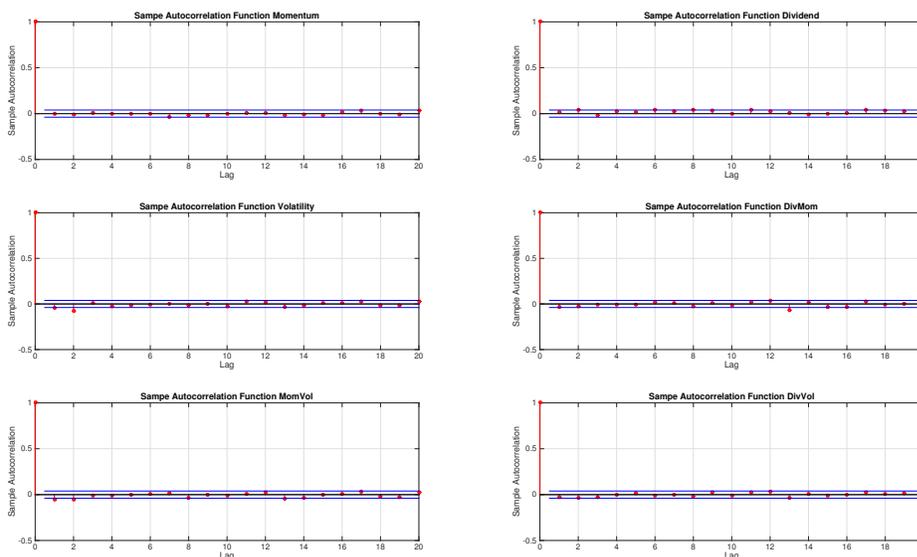


Figure 3.5: *Autocorrelation function for the residuals of each index. From the upper left: Momentum, Dividend, Volatility, Dividend-Momentum, Momentum-Volatility, Dividend-Volatility.*

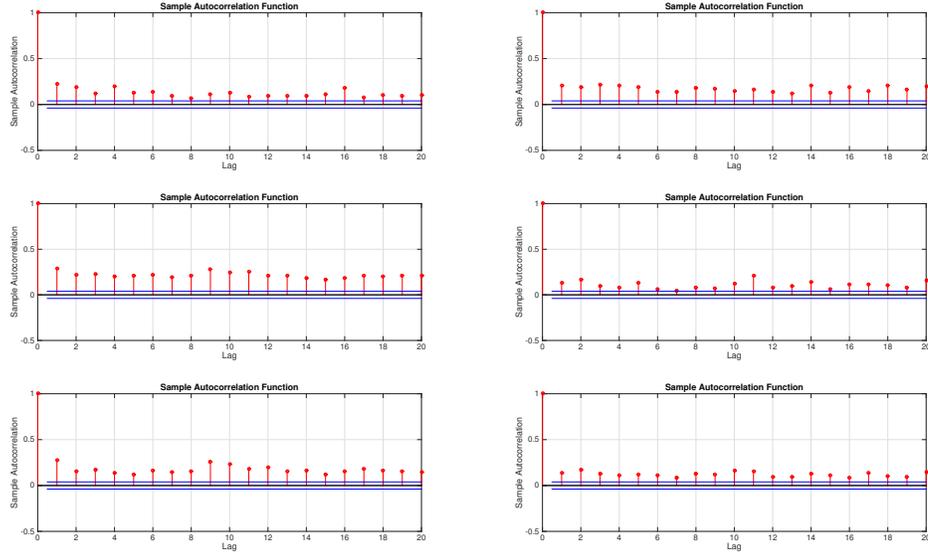


Figure 3.6: *Autocorrelation function for the squared residuals of each index. **From the upper left:** Momentum, Dividend, Volatility, Dividend-Momentum, Momentum-Volatility, Dividend-Volatility.*

To further test for dependence in the data, the Ljung-Box test is applied to the residuals and the squared residuals and Engle’s ARCH test is applied to the residuals to test for heteroscedasticity. The Ljung-Box test for the squared residuals rejects the null-hypothesis that there is no autocorrelation with a p-value of zero for all indices for lags one to eight, which coincides with the result from Figure 3.6. The Ljung-Box test on the residuals however rejects the null hypothesis for the volatility and momentum-volatility index.

These properties, no autocorrelation in the residuals but a significant autocorrelation in the squared residuals, indicates that there is heteroscedasticity in the data [3]. To test for heteroscedasticity, Engle’s test for residual heteroscedasticity is applied for lags varying from 1 to 10. The test rejects the hypothesis that there is no heteroscedasticity in the residuals with a p-value of zero for all lags and indices.

3.2 Investment Strategy

The signal-value will be incorporated into two portfolio strategies and the equally weighted portfolio will be used as a benchmark. The different strategies are used to demonstrate the impact of the signal-value and the weights will be re-evaluated monthly.

Let the portfolio weights be denoted by the vector $\omega_{\mathbf{t}}$, where

$$\omega_{\mathbf{t}} = (\omega_{1,t}, \dots, \omega_{N,t}), \quad (3.1)$$

and N is the number of assets. The constraint $\omega_i \geq 0$ is imposed so that no short selling is allowed. Then the value of the portfolio at time $t = T$ can be expressed as:

$$V_T = V_0 \prod_{t=1}^T \omega_{\mathbf{t}} r_t \quad (3.2)$$

where r_t is the vector of returns at time t .

The first approach when having found a signal-value is to incorporate this signal-value in the $1/N$ portfolio. That is, if the signal-value indicates that one or more of the indices are going to perform poorly, all of the wealth will be put into the index appearing to perform best in a crisis.

The second approach is to minimize the portfolio variance while setting a lower bound on the expected return. This is a quadratic optimization problem that can be expressed as:

$$\begin{aligned} \underset{\omega_{\mathbf{t}}}{\text{minimize}} \quad & \omega_{\mathbf{t}}^T \mathbf{H}_{\mathbf{t}+1} \omega_{\mathbf{t}} \\ \text{subject to} \quad & \omega_{\mathbf{t}} \mu_{\mathbf{t}+1} \geq \mu_t^* \\ & \omega_{\mathbf{t}}^T \mathbf{1} = 1, \\ & w_{i,t} \geq 0, \quad i = 1, \dots, N \end{aligned} \quad (3.3)$$

where $\mathbf{H}_{\mathbf{t}+1}$ is the covariance matrix and $\mu_{\mathbf{t}+1}$ is the vector of expected values of each index at time $t = t + 1$. The constant μ_t^* is chosen as the expected return of the benchmark portfolio, $\mathbf{r}_{\mathbf{t}} \mathbf{1}^T / N$, according to [17].

This is a convex optimization problem, which can be seen by rewriting the constraint $\omega_{\mathbf{t}} \mu_{\mathbf{t}+1} \geq \mu_t^*$ with $-\omega_{\mathbf{t}} \mu_{\mathbf{t}+1} \leq -\mu_t^*$ and $w_{i,t} \geq 0$ with $-w_{i,t} \leq 0$. This approach will be tested with and without the influence of the indicators. The indicators will be incorporated in the optimization in the same way as for the naïve approach.

This problem can be solved by using the matlab function `fmincon()` which accepts both linear and non linear equality and inequality constraints. The `fmincon()` function uses the interior-point algorithm by default, which can handle both large-scale sparse problems and small dense problems. The interior-point algorithm as used by the `fmincon()` function is described in Section 2.6.

Chapter 4

Results and Discussion

4.1 Model Selection

From the analysis of the autocorrelation and the results from Engle's test for heteroscedasticity, a GARCH(p,q) model seem appropriate to model the daily returns of the outperformance. The order p and q can be estimated by Akaike's information criterion, Bayesian information criterion, or other criteria, see for example [24], [16]. In this thesis, the returns are assumed to follow a GARCH(1,1) process, which has proven to suffice in most applications, see for example [6], to reduce the number of parameters to be estimated.

Three distributions is tested for the model; the normal distribution, the Student's t distribution and the skew-t distribution. The approximated parameters of each distribution can be seen in Table 4.1. For the normal and student's t distribution, the parameters are typical for those found for daily returns with the GARCH lag close to one and the ARCH lag and constant close to zero. The sum of the parameters $\alpha_{i,1}$ and $\beta_{i,1}$ for each index is close to one, indicating high volatility persistence for the normal and Student's t distributions. The sum of θ_1 and θ_2 are close to one for all distributions, especially for the normal. This indicates a slow rate of mean-reversions in the correlations [13].

	Normal Distribution	Student's t distribution	Skew t- Distribution
$\alpha_{1,0}$	0.0176	0.0147	1.1357
$\alpha_{2,0}$	0.0104	0.0088	2.0042
$\alpha_{3,0}$	0.0102	0.0069	1.4646
$\alpha_{4,0}$	0.0074	0.0062	0.1273
$\alpha_{5,0}$	0.0098	0.0073	3.6062
$\alpha_{6,0}$	0.0153	0.0063	1.8625
$\alpha_{1,1}$	0.8684	0.9061	0.2227
$\alpha_{2,1}$	0.9107	0.9390	0.2553
$\alpha_{3,1}$	0.8988	0.9407	0.2856
$\alpha_{4,1}$	0.9194	0.9433	0.3111
$\alpha_{5,1}$	0.9020	0.9533	0.2391
$\alpha_{6,1}$	0.8639	0.9483	0.2987
$\beta_{1,1}$	0.1306	0.0929	0.2017
$\beta_{2,1}$	0.0883	0.0582	0.3763
$\beta_{3,1}$	0.1002	0.0583	0.3613
$\beta_{4,1}$	0.0796	0.0557	0.3294
$\beta_{5,1}$	0.0947	0.0637	0.3794
$\beta_{6,1}$	0.1279	0.0488	0.3399
θ_1	0.4762	0.2833	0.5560
θ_2	0.5016	0.6717	0.2213
ν	-	6.9030	2.1898
ξ_1	-	-	16.1236
ξ_2	-	-	0.1699
ξ_3	-	-	16.5962
ξ_4	-	-	20.1216
ξ_5	-	-	0.1541
ξ_6	-	-	22.4434

Table 4.1: *Estimated parameters for TVC-GARCH(1,1) models for normal distribution, Student's t -distribution and skew t -distribution.*

When the model has been fitted to the data, the GARCH residuals should be an IID process. To confirm this, the autocorrelation and squared autocorrelation function of ε_t are investigated. As can be seen in Figure 4.1-4.3, the squared autocorrelations of the GARCH residuals are zero for the normal and the student's t distribution, meaning that they have successfully removed the GARCH effects. The model with the skew t -distribution however has not succeeded to remove the GARCH effects in the series.

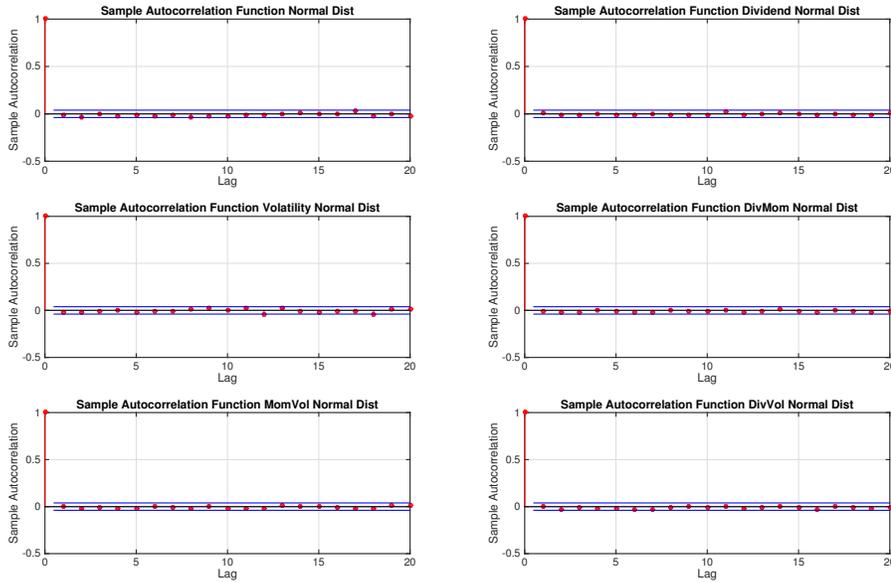


Figure 4.1: *The squared autocorrelations for the garch residuals of each index when assuming a normal distribution. **From the upper left:** Momentum, Dividend, Volatility, Dividend-Momentum, Momentum-Volatility, Dividend-Volatility.*

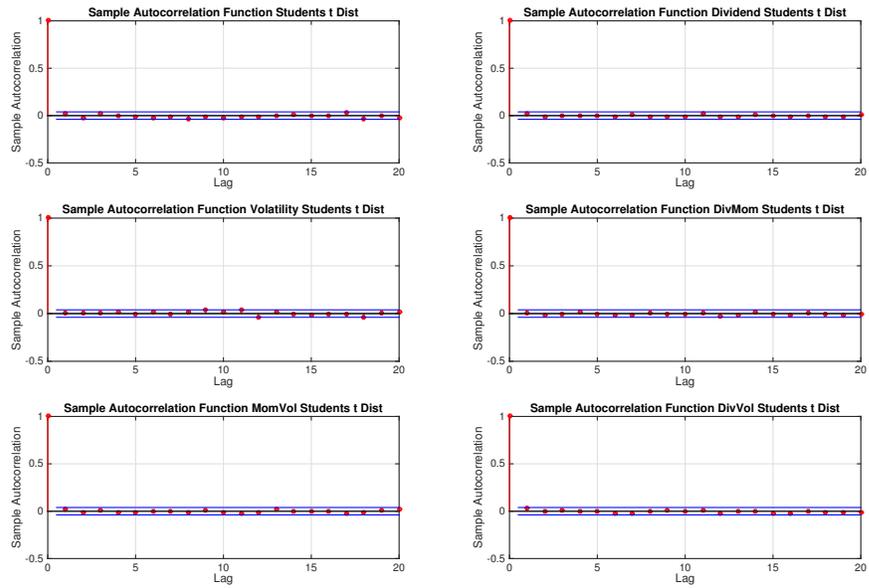


Figure 4.2: *The squared autocorrelations for the garch residuals of each index when assuming student's t -distribution. **From the upper left:** Momentum, Dividend, Volatility, Dividend-Momentum, Momentum-Volatility, Dividend-Volatility.*

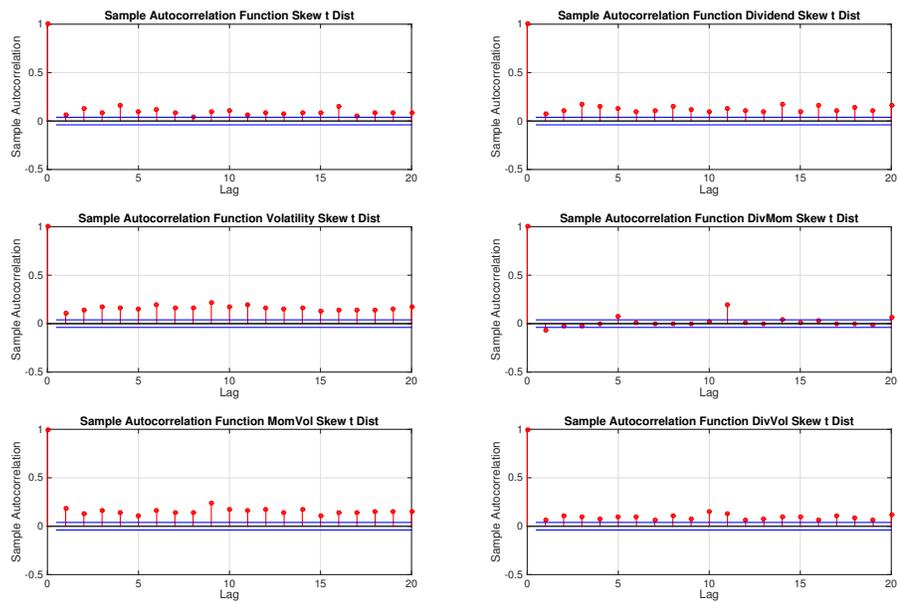


Figure 4.3: *The squared autocorrelations for the garch residuals of each index when assuming a skew t -distribution. **From the upper left:** Momentum, Dividend, Volatility, Dividend-Momentum, Momentum-Volatility, Dividend-Volatility.*

In Figure 4.4-4.6, the QQ-plots of the sample quantiles against the quantiles of the fitted distributions are shown. If the sample is from the distribution tested in the QQ-plot, the plot will be linear. In Figure 4.4, the Gaussian error distribution is plotted against the sample. It can be seen that the sample data is more heavy tailed than the Gaussian distribution. This is true for both the left and the right tail, hence the data might not be skewed. In Figure 4.5, it can be seen that the Student's t distribution seem to capture both of the tails and provide a good fit for the sample data. The skewed distribution, as can be seen in Figure 4.6, highly underestimated the left tail while providing a reasonably good fit for the right tail. From these figures, it can be seen that the Student's t -distribution with $\nu = 6.9$ degrees of freedom seems to be the best fit to the data.

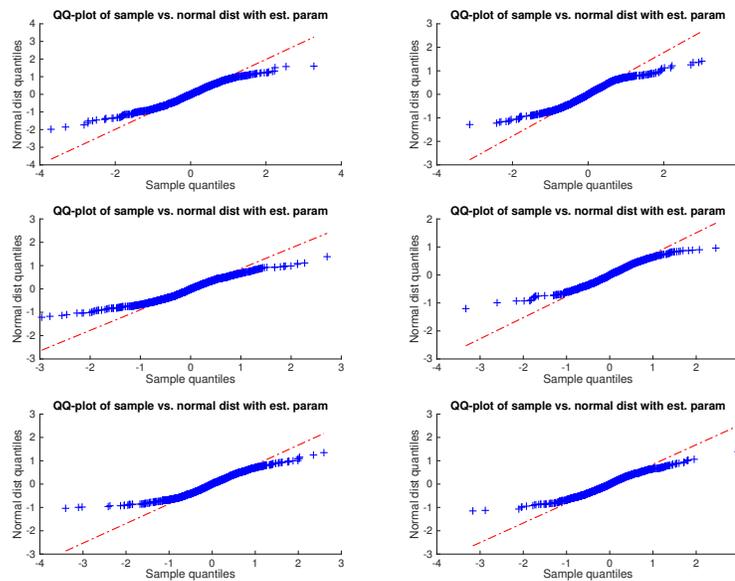


Figure 4.4: *QQ-plot of the sample against the normal distribution. From the upper left: Momentum, Dividend, Volatility, Dividend-Momentum, Momentum-Volatility, Dividend-Volatility.*

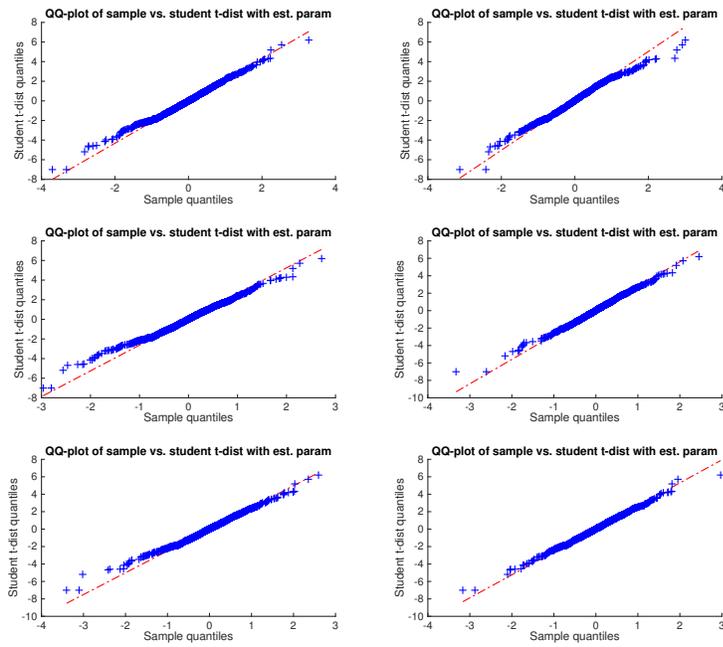


Figure 4.5: *QQ-plot of the sample against the Student's t -distribution. From the upper left: Momentum, Dividend, Volatility, Dividend-Momentum, Momentum-Volatility, Dividend-Volatility.*

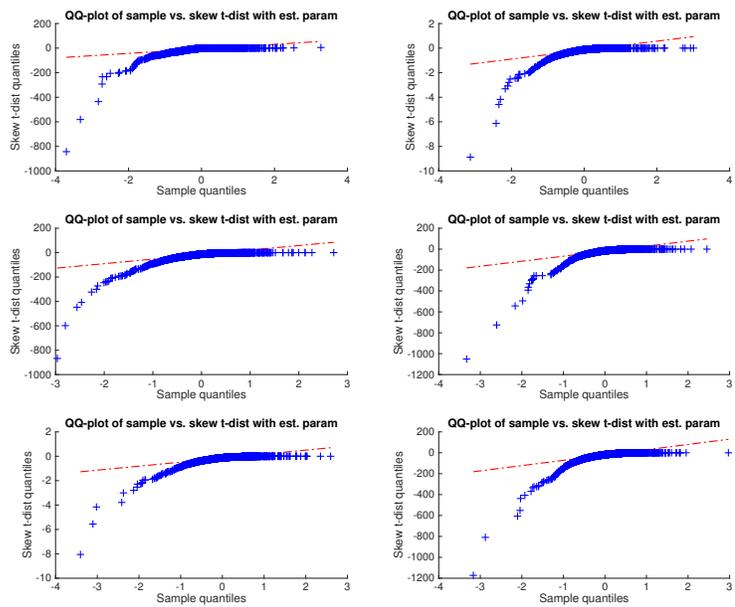


Figure 4.6: *QQ-plot of the sample against the skew t-distribution. From the upper left: Momentum, Dividend, Volatility, Dividend-Momentum, Momentum-Volatility, Dividend-Volatility.*

4.2 Forecasting the Correlation

When the Student's t distribution is determined to best fit the data, the conditional variance and correlation can be calculated. 220 values are set aside for the forecast and the result of the inferred variance and the forecast for the conditional variance can be seen in Figure 4.7. In Figure 4.8, the forecast of the conditional variance when using 3000 values is shown. It can be seen that the forecasted conditional variance converges to the unconditional variance for all indices. The convergence is slow for all indices, approximately 1000 steps for the Dividend and Dividend-Volatility index and 3000 steps for the others. This is due to the large value of $(\alpha_i + \beta_i)$, which is the exponential rate of decay of the memory of D_{t+k} . The values of (α_i, β_i) for the Dividend and Dividend-Volatility index are smaller than those of the other indices.

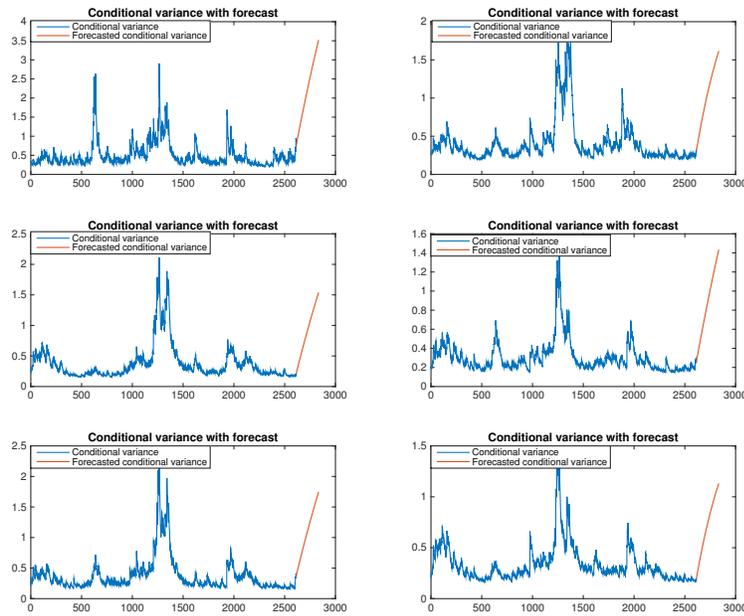


Figure 4.7: The inferred and forecasted conditional variance **From the upper left: Momentum, Dividend, Volatility, Dividend-Momentum, Momentum-Volatility, Dividend-Volatility.**

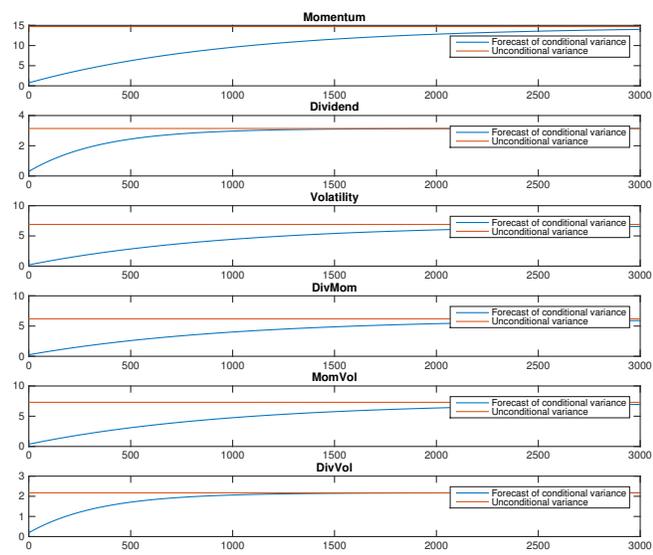


Figure 4.8: *The forecasted conditional variance as the number of forecast steps increases.*

The forecasts for the correlation between the momentum and dividend index is shown in Figure 4.9, and it can be seen that the forecasted correlation converges to the correlation of ε_i as the forecast step k increases. This holds for all indices. The forecasted conditional correlation matrix at $k=220$ is:

$$\Gamma_{t+220} = \begin{bmatrix} 1.0000 & 0.3797 & 0.3894 & 0.5845 & 0.6587 & 0.3961 \\ 0.3797 & 1.0000 & 0.6157 & 0.7114 & 0.5593 & 0.7156 \\ 0.3894 & 0.6157 & 1.0000 & 0.6271 & 0.6898 & 0.8328 \\ 0.5845 & 0.7114 & 0.6271 & 1.0000 & 0.7674 & 0.7340 \\ 0.6587 & 0.5593 & 0.6898 & 0.7674 & 1.0000 & 0.6781 \\ 0.3961 & 0.7156 & 0.8328 & 0.7340 & 0.6781 & 1.0000 \end{bmatrix} \quad (4.1)$$

whereas the unconditional correlation matrix of the GARCH residuals is:

$$\Gamma_\varepsilon = \begin{bmatrix} 1.0000 & 0.3797 & 0.3894 & 0.5845 & 0.6587 & 0.3961 \\ 0.3797 & 1.0000 & 0.6157 & 0.7114 & 0.5593 & 0.7156 \\ 0.3894 & 0.6157 & 1.0000 & 0.6271 & 0.6898 & 0.8328 \\ 0.5845 & 0.7114 & 0.6271 & 1.0000 & 0.7674 & 0.7340 \\ 0.6587 & 0.5593 & 0.6898 & 0.7674 & 1.0000 & 0.6781 \\ 0.3961 & 0.7156 & 0.8328 & 0.7340 & 0.6781 & 1.0000 \end{bmatrix}. \quad (4.2)$$

That is, the conditional correlation matrix has, as expected, converged to the correlation matrix of the GARCH residuals. The convergence to the unconditional correlation takes approximately 250 steps, which depends on the value of $(\theta_1 + \theta_2)$, the rate of decay of the memory of Γ_{t+k} .

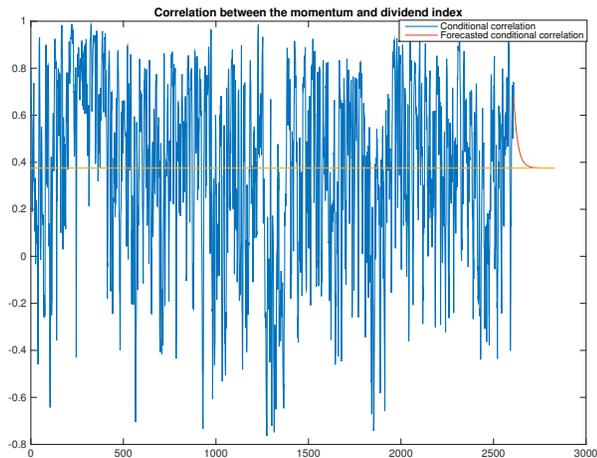


Figure 4.9: *The forecasted conditional correlation between the momentum and dividend index.*

4.3 Portfolio Performance

The portfolio performance over the whole time period, measured as the portfolio value at each time point t , with the forecasted values marked with red can be seen in Figure 4.10 for the four different portfolios. In Figure 4.11, the cumulative sum of the portfolio returns is shown for the whole time period, with the forecasted values marked with red. It can be seen that the mean-variance portfolio based on the indicators accumulates the highest return while the naïve strategy accumulates the lowest return. This coincides with the mean return of the portfolios, found in Table 4.3, where the mean-variance portfolio with indicators has the highest mean return for the in-sample period. It also has the highest Sharpe ratio in the in-sample period. The p-values shown in Table 4.3 are the bootstrap p-values (assuming IID series) from the null hypothesis that the Sharpe ratio is equal to that of the $1/N$ portfolio [18]. In the in-sample period, it can be seen that the portfolios based on the indicators have Sharpe ratios higher than that of the $1/N$ portfolio, and that they are statistically different at a 10 % significance level. The mean-variance portfolio without indicators however performs statistically as well as the $1/N$ portfolio in the in-sample period.

In the out-of-sample period however, the mean-variance has the highest Sharpe ratio. Though when looking at the p-values, there is no statistical difference between the performances of the portfolios compared to the $1/N$ portfolio. This coincides with the findings in [9], that states that there is no model that delivers an out-of-sample Sharpe ratio higher than that of the $1/N$ portfolio.

The mean-variance portfolios are exposed to estimation risk from the fact that \mathbf{H}_{t+1} and μ_{t+1} are estimated quantities. The portfolios based on the indicators, that have the highest standard deviation, are also exposed to estimation risk due to the linear combination of indicators used to switch strategies.

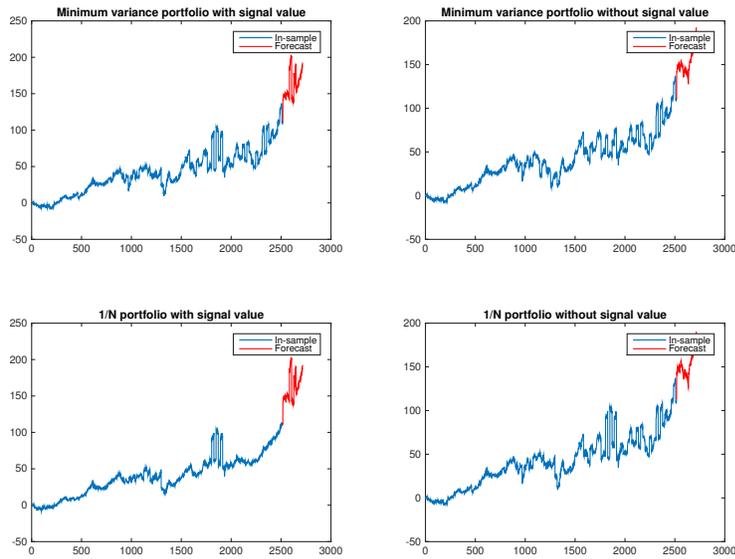


Figure 4.10: *The performance of all portfolios, with the forecasted values in red. **From the upper left:** Mean-variance portfolio with indicators, mean-variance portfolio without indicators, 1/N portfolio with indicators, 1/N portfolio without indicators.*

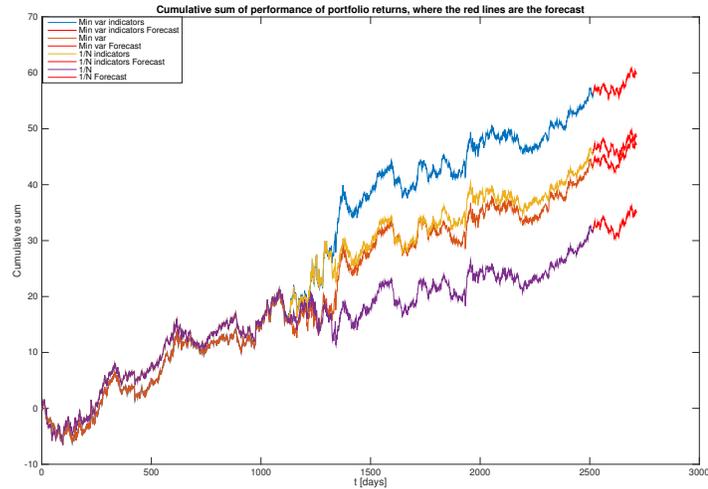


Figure 4.11: *The cumulative sum of returns of all portfolios, with the forecasted values in red. **From the upper left:** Mean-variance portfolio with indicators, mean-variance portfolio without indicators, 1/N portfolio with indicators, 1/N portfolio without indicators.*

Statistics of the portfolios out-of-sample performance	Mean	Standard Deviation	Sharpe Ratio	p-Value
1/N	0.0166	0.2779	0.0599	1
1/N with indicators	0.0149	0.3003	0.0496	0.8104
Mean-Variance	0.0192	0.2488	0.0773	0.4232
Mean-Variance with indicators	0.0189	0.2809	0.0675	0.4930
Statistics of the portfolios in-sample performance				
1/N	0.0126	0.4211	0.0300	1
1/N with indicators	0.0182	0.4452	0.0408	0.0660
Mean-Variance	0.0173	0.4177	0.0413	0.1318
Mean-Variance with indicators	0.0223	0.4402	0.0507	0.0218

Table 4.2: *Statistics of the four portfolios for both in- and out-of-sample periods.*

Chapter 5

Conclusions and Remarks

The TVC-GARCH(1,1) model with Student's t -distributed errors is chosen for the data. The Student's t -distribution provides a good fit for the in-sample data and the order of the lags are chosen to minimize the number of parameters estimated and the mean is approximated as the sample mean. In order to provide a better fit a conditional mean model could have been implemented, for example an AR(1). All of the parameters are determined by maximum-likelihood estimation, and this is done in one step in order to improve the estimation result. To further improve the estimation result, it might have been beneficial to consider marginal models such as EGARCH or GJR-GARCH that can capture asymmetric conditional variances.

A linear combination of the indicators Ifo, GDP, CPI, Labour and PMI are regressed on the dividend index and used as a signal value for when the indices are about to drop. This signal value could be improved by usage of more or different indicators, and a regression other than a linear might provide a better explanation for the index movements. All of the parameters in the regression are however significant at 5 %, and provides a good fit to the data.

The portfolios are set up by implementing the signal value in the 1/N and the mean-variance portfolio. The main purpose of the portfolios were to demonstrate the impact of the signal-value and there exists many other strategies that could have been implemented, such as maximization of the Sharpe ratio, the trade-off problem or the minimum-variance portfolio. In the in-sample period, the mean-variance portfolio with the signal-value outperforms the other portfolios and its Sharpe ratio is significantly higher than that of the 1/N portfolio. Thus the linear combination of indicators seem to explain the behaviour of the indices. In the out-of-sample period however, the portfolios depending on the indicators have lower mean returns than the portfolios that are independent of the signal-value. The Sharpe ratios of the portfolios are not statistically different from that of the 1/N portfolio, hence

these strategies are not optimal to implement.

Bibliography

- [1] Arnott, Robert D., Hsu, Jason & Moore, Philip
Fundamental Indexation Financial Analysts Journal, Vol. 61 (2005), pp. 83-99.
- [2] Bauwens, Luc & Laurent, Sébastien
A New Class of Multivariate Skew Densities, with Application to Generalized Autoregressive Conditional Heteroscedasticity Models.
Journal of Business & Economic Statistics, Vol. 23 (2005), pp. 346-354.
- [3] Brockwell, Peter J. & Davis, Richard A.
Introduction to Time Series and Forecasting, Second Edition.
Springer-Verlag New York Inc., (2002)
- [4] Bollerslev, Tim
Generalized Autoregressive Conditional Heteroskedasticity.
Journal of Econometrics, Vol 31 (1986), pp. 307-327
- [5] Bollerslev, Tim
Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model.
The Review of Economics and Statistics, Vol. 72 (1990), pp.498-505
- [6] Bollerslev, Tim., Chou, Ray Y. & Kroner, Kenneth F.
ARCH modeling in finance.
Journal of Econometrics, Vol 52 (1992), pp. 5-59
- [7] Campbell, Harvey R.
Large-cap, Definition.
Available at:<http://www.nasdaq.com/investing/glossary/l/large-cap>
- [8] Campbell, Harvey R.
Small-capitalization (small cap) stocks, Definition.
Available at: <http://www.nasdaq.com/investing/glossary/s/small-cap-stocks>

- [9] DeMiguel, Victor & Garlappi, Lorenzo & Uppal, Raman
Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy.
The Review of Financial Studies, V. 22 (2009), pp. 1915-1953
- [10] Engle, Robert
Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation.
Econometrica, Vol. 50, (1982), pp. 987-1002
- [11] Engle, Robert
Dynamic Conditional Correlation - A Simple Class of Multivariate GARCH Models.
Journal of Business & Economic Statistics, Vol. 20 (2002), pp. 339-350
- [12] Engle, Robert F., & Sheppard, Kevin
Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH.
National Bureau of Economic Research (2001)
- [13] García-Álvarez, Luis, & Luger, Richard
Dynamic Correlations, Estimation Risk, and Portfolio Management During the Financial Crisis. CEMFI, Working Papers.
- [14] IFAU - Institutet för arbetsmarkands- och utbildningspolitisk utvärdering
Arbetskraftsdeltagande, sysselsättning och arbetslöshet.
Available at: <http://www.ifau.se/sv/Forskningsomraden/Arbetskraftsdeltagande-sysselsattning-och-arbetsloshet/>
- [15] Ifo Institute
Calculating the Ifo Business Climate
Available at: <http://www.cesifo-group.de/ifoHome/facts/Survey-Results/Business-Climates/Calculating-the-Ifo-Business-Climates.html>
- [16] Javed, Farrukh & Mantalos, Panagiotis
GARCH-type Models and Performance of Information Criteria.
Communications in Statistics-Simulation and Computation, Vol 42 (2013), pp.1917-1933
- [17] Kirby, Chris & Ostdiek, Barbara
It's All in the Timing: Simple Active Portfolio Strategies that Outperform Naïve Diversification.
Journal of Financial and Quantitative Analysis, Vol 47 (2012) pp.437-467
- [18] Ledoit, Oliver & Wolf, Michael
Robust performance hypothesis testing with the Sharpe ratio.
Journal of Empirical Finance, V. 15 (2008), pp. 850-859

- [19] Ljung, G. M. & Box, G. E. P.
On a measure of lack of fit in time series models.
Biometrika, Vol. 65 (1978), pp. 297-303
- [20] Markit
PMI.
Available at: <https://www.markit.com/product/pmi>
- [21] Markowitz, Harry *Portfolio Selection.*
The Journal of Finance, Vol. 7 (1952), pp. 77-91.
- [22] Mathworks
fmincon.
Available at: <http://se.mathworks.com/help/optim/ug/fmincon.html?refresh=true>
- [23] Mathworks
Constrained Nonlinear Optimization Algorithms.
Available at: <http://se.mathworks.com/help/optim/ug/constrained-nonlinear-optimization-algorithms.html>
- [24] Mitchell, Heather & McKenzie, Michael D.
GARCH model selection criteria.
Quantitative Finance, Volume 3 (2003) pp.262-284
- [25] The Nasdaq OMX Group, Inc.
Sweden Smart Beta Indexes.
Available at: <https://indexes.nasdaqomx.com/Home/SwedenSmartBeta>
- [26] The Nasdaq OMX Group, Inc.
Rules for the Construction and Maintenance of the Nasdaq Nordea SmartBeta Index Family.
Available at: https://indexes.nasdaqomx.com/docs/Methodology_NQNDSmartBeta.pdf
- [27] Petch, Kim
Investing Strategies & Styles - Are You an Alpha or Beta Investor?
Available at: <http://www.moneycrashers.com/investing-strategies-styles-beta-alpha-investment/>
- [28] Statistiska Centralbyrån
Konsumentprisindex (KPI).
Available at: http://www.scb.se/sv_/Vara-tjanster/Index/Konsumentpriser/Konsumentprisindex-KPI/

- [29] Statistiska Centralbyrån
Bruttonationalprodukt (BNP).
Available at: http://www.scb.se/sv_/Hitta-statistik/Statistik-efter-amne/Nationalrakenskaper/Nationalrakenskaper/Nationalrakenskaper-kvartals-och-arsberakningar/22909/Begreppsforklaringar-inom-nationalrakenskaperna/#Bruttonationalprodukt
- [30] Swedbank
Mål och metodik för Inköpschefsindex.
Available at: <https://www.swedbank.se/om-swedbank/analyser/swedbank-makroanalys/inkopschefsindex-industri/inkopschefsindex,-mal-och-metodik/index.htm>
- [31] Tsay, Ruey S.
Conditional Heteroscedastic Models
Analysis of Financial Time Series, Third Edition (2010), Third Edition, John Wiley & Sons, Inc., Hoboken, NJ, USA.
- [32] Tse, Y.K and Tsui, Albert K.C
A Multivariate GARCH Model with Time-Varying Correlations
Journal of Business & Economic Statistics, Vol. 20 (2002), pp. 351-362

Appendix A

Regression statistics

	Momentum	Dividend	Volatility	DivMom	MomVol	DivVol
Intercept	1198.8626 (912.73)	2685.7529** (743.54)	1114.6975 (760.1)	1421.9560 (810.66)	243.3927 (901.82)	1509.0445** (483.05)
Ifo	-3.7906** (0.51069)	-2.7199** (0.43841)	-3.3051** (0.42529)	-3.3973** (0.45359)	-3.1680** (0.50459)	-1.8180** (0.27028)
GDP	15.0581** (1.3452)	12.0365** (1.1548)	12.9031** (1.1202)	12.7862** (1.1948)	11.6369** (1.3291)	7.6545** (0.71191)
CPI	-9.6754** (1.064)	-4.6953** (0.91343)	-6.4732** (0.88611)	-5.7143** (0.94506)	-2.3701* (1.0513)	-3.8993** (0.56313)
Labour	-15.0047 (9.4505)	-33.2688** (8.1129)	15.6152* (-7.8702)	-19.6777* (8.3937)	-10.3373 (9.3376)	-17.9087** (5.0015)
PMI	1.2080** (0.28953)	1.9532** (0.24855)	1.7336** (0.24111)	2.0891** (0.25715)	2.1898** (0.28607)	1.0741** (0.15323)
Gov 5Y	0.1714 (0.11093)	-0.1737 (0.095225)	-0.5213** (0.092377)	-0.3648** (0.098522)	-0.4624** (0.1096)	-0.2463** (0.058706)
Root mean Squared error	16.7	14.3	13.9	14.8	16.5	8.82

Table A.1: Statistics for regression with all indicators, estimates marked with * are significant at a 5 % level, estimates marked with ** are significant at a 1 % level. The standard errors are written in the parenthesis.

	Momentum	Dividend	Volatility	DivMom	MomVol	DivVol
Intercept	-	3386.6** (689.49)	-	-	-	1509.0445** (483.05)
Ifo	-4.1141** (0.4962)	-3.0563** (0.4016)	-3.1534** (0.4145)	-3.2038** (0.44376)	-2.4163** (0.72813)	-1.8180** (0.27028)
GDP	14.616** (1.1064)	12.936** (1.0544)	11.843** (0.85974)	11.434** (0.92043)	11.042** (1.5204)	7.6545** (0.71191)
CPI	-11.578** (0.89642)	-4.595** (0.92058)	-5.9242** (0.80699)	-5.014** (0.86396)	-8.4801* (1.2113)	-3.8993** (0.56313)
Labour	-	-41.193** (6.9182)	-4.1* (-0.53639)	-4.9885* (0.57426)	-	-17.9087** (5.0015)
PMI	1.1745** (0.28327)	2.0238** (0.24789)	1.18422** (0.23054)	2.2277** (0.24682)	1.1771** (0.38801)	1.0741** (0.15323)
Gov 5Y	-	-	-0.58769** (0.080897)	-0.44948** (0.086608)	-1.3998** (0.089161)	-0.2463** (0.058706)
Root mean Squared error	18.4	14.4	13.9	14.9	24.7	8.82

Table A.2: Statistics for regression with the significant indicators from Table A.2, estimates marked with * are significant at a 5 % level, estimates marked with ** are significant at a 1 % level. The standard errors are written in the parenthesis.