

Evaluating Dynamic Covariance Matrix Forecasting and Portfolio Optimization

By Lars Hegnes Sendstad and Dag Martin Holten

Advisor: Jonas Andersson

Master thesis within the main profile Economic Analysis/Finance

This thesis was written as a part of the Master of Science in Economics and Business Administration program – Economic Analysis/Finance. Neither the institution, nor the advisor is responsible for the theories and methods used, or the results and conclusions drawn, through the approval of this thesis.

Abstract

In this thesis we have evaluated the covariance forecasting ability of the simple moving average, the exponential moving average and the dynamic conditional correlation models. Overall we found that a dynamic portfolio can gain significant improvements by implementing a multivariate GARCH forecast. We further divided the global investment universe into sectors and regions in order to investigate the relative portfolio performance of several asset allocation strategies with both variance and conditional value at risk as a risk measure. We found that the choice of risk measure does not seem to heavily impact the asset allocation. As comparison to the dynamic portfolios we added regional/sector portfolios which were rebalanced after a 3% threshold rule. The regional portfolio was constructed to mimic the current strategy of the Norwegian Pension Fund Global. The max Sharpe portfolio for regions had the highest risk adjusted return, but suffered from a very high turnover. After being modified however, this strategy turned out to be superior even after transaction costs were imposed.

Contents

CONTENTS	3
1. INTRODUCTION	6
2. COVARIANCE MATRIX FORECASTING METHODS.....	10
2.1 SIMPLE MOVING AVERAGE	10
2.2 EWMA	10
2.3 UNIVARIATE GARCH MODEL	11
2.4 DCC-GARCH	12
2.4.1 <i>The DCC-GARCH Model</i>	12
2.4.2 <i>Estimation of the DCC(1,1) Model</i>	13
2.4.3 <i>Forecasting the Covariance Matrix</i>	15
3. MEASURING RISK	16
3.1 VARIANCE.....	16
3.2 VALUE AT RISK.....	16
3.3 CONDITIONAL VALUE AT RISK.....	17
3.3.1 <i>CVaR under Normality:</i>	19
3.4 THE MODIFIED CVAR ESTIMATOR	20
3.4.1 <i>The Cornish-Fisher Expansion</i>	20
4. METHODOLOGY FOR COVARIANCE AND RETURN EVALUATION	22
4.1 COVARIANCE FORECAST EVALUATION	22
4.2 EXPECTED RETURN	24
5. PORTFOLIO OPTIMIZATION THEORY	25
5.1 MINIMUM VARIANCE	25
5.2 MAXIMUM SHARPE	26

5.3	MINIMIZING VARIANCE AND CVAR WITH A RETURN TARGET	26
5.4	RISK BUDGETS WITH CVAR AND VARIANCE	27
5.5	REGIONAL AND SECTOR REBALANCING STRATEGIES	28
6.	DESCRIPTION OF THE DATASET	30
6.1	REGIONS.....	30
6.2	SECTORS.....	32
7.	RESULTS FROM THE COVARIANCE AND RETURN EVALUATION.....	36
7.1	FITTING MODELS.....	36
7.2	EXPECTED RETURN	37
7.3	EVALUATING THE PERFORMANCE OF DIFFERENT COVARIANCE FORECASTSING METHODS ..	38
7.4	STATIONARITY	39
8.	PORTFOLIO OPTIMIZATION RESULTS.....	40
8.1	THE MINIMUM VARIANCE AND MAX SHARPE PORTFOLIOS	40
8.2	REGIONAL PORTFOLIOS WITH ADDITIONAL CONSTRAINTS	47
8.3	SECTOR RESULTS WITH ADDITIONAL CONSTRAINTS.....	51
9.	THE MODIFIED REGIONAL MAX SHARPE PORTFOLIO	55
10.	CONCLUSION	59
11.	REFERENCES	60
12.	APPENDIX.....	64
12.1	APPENDIX A: DATA CLEANING	64
12.2	APPENDIX B: STATIONARITY	65
12.3	APPENDIX C: DIAGNOSTIC PLOTS	67
12.3.1	<i>Regions</i>	67
12.3.2	<i>Sector</i>	73
12.4	APPENDIX D: DERIVING VAR	79

12.5	APPENDIX E: DIFFERENTIAL EVOLUTION ALGORITHM	81
------	--	----

1. Introduction

Portfolio diversification is typically achieved through investing in different asset classes, or different assets that are thought to have low or negative correlation. This is a strategy that has strong empirical evidence and theoretical justification, but an investor must be aware that the correlation between assets varies over time, which implies that the degree of portfolio diversification attainable in a given portfolio will be time dependent. A number of studies find that correlation between equity returns increase during bear markets, and decrease during bull markets (Ang and Bekaert (2001), Das and Uppal (2001), and Longin and Solnik (2001)). Another well-known stylized fact is volatility clustering, meaning that large deviation tends to be followed by large deviation i.e. autocorrelation in variance. In addition negative returns tend to be followed by larger increases in the volatility than positive returns. This is known as the “leverage effect”, however research suggests that the leverage effect observed in financial time series is not fully explained by the firm’s leverage. See Hens and Steude (2009) and Figlewski and Wang (2000)

Modelling volatility in financial time series has of course been the object of much attention given stylized facts as those mentioned above. “The presence of volatility clusters suggests that it may be more efficient to use only the most recent observations to forecast volatility, or perhaps assign higher weight to the most recent observations” Danielsson (2011). The first conditional volatility model introduced was the Autoregressive Conditional Heteroskedasticity (ARCH) model by Engle (1982). Subsequently, numerous variants and extensions of ARCH models have been proposed, as for example the generalized Autoregressive Conditional Heteroskedasticity (GARCH) model. For a review of volatility models see Lundbergh and Teräsvirta (2002). While modelling volatility of univariate returns is well understood, understanding the co-movements of financial assets is a much more complex problem. Construction of a variance optimized portfolio requires a forecast of the covariance matrix. Such applications entail estimation and forecasting of large covariance matrices, potentially with thousands of assets. The search for reliable estimates of correlations between financial assets has been the subject of a lot of research and simple methods such as rolling historical correlations and exponential smoothing of historical returns are widely used. The univariate conditional volatility models have been extended to multivariate GARCH (MGARCH) models. But the multivariate GARCH models quickly get

too complex as the number of assets increases and are seldom estimated for more than five assets. The dynamic conditional correlation GARCH (DCC) model introduced by Engle (2001) has the flexibility of univariate GARCH models but not the complexity of conventional multivariate GARCH models. These models are estimated in two steps, the first is a series of univariate GARCH estimates and the second the correlation estimate. This method has a clear computational advantage over other multivariate GARCH models in that the number of parameters to be estimated in the second process is independent of the number of series to be correlated (Engle (2002)). Thus potentially very large correlation matrices can be estimated. The simple parameterization of the model assumes the same dynamic correlation process, and can therefore be seen as a weakness.

Markowitz (1952) introduced the mean-variance risk management framework. This is optimal if returns are normally distributed or the investor has a quadratic utility function. In the late 1980s Value at Risk (VaR) emerged as a distinct concept and has become a widely used and popular measure of risk (J.P. Morgan (1994)). The popularity of VaR is mostly related to its simple and easy to understand representation of high losses. VaR can be efficiently estimated when the underlying risk factors are normally distributed. However, for non-normal distributions, VaR may have undesirable properties (e.g Artzner et al. (1999)). Such a property is its lack of sub additivity, meaning that VaR of a portfolio with several instruments may be greater than the sum of the individual VaRs. Also Mauser and Rosen (1999) and many more have showed that VaR can be problematic in determining an optimal mix of assets, since it can exhibit multiple local extrema. Because of these weaknesses we will apply another percentile risk measure, Conditional Value at Risk (CVaR) instead of VaR. For continuous distributions, CVaR is defined as the conditional expected loss under the condition that it exceeds VaR, see Rockafellar and Uryasev (2000). CVaR has more attractive properties than VaR, because it is sub-additive and convex (Rockafellar and Uryasev, 2000). Moreover, CVaR is a coherent measure of risk in the sense of Artzner et al. (1999). This measure is also able to incorporate higher moments of the return distribution; without placing any specific assumption on risk aversion.

The goal of this thesis is two folded; first we will evaluate the simple moving average, exponentially weighted moving average (EWMA) and DCC covariance matrix forecasting methods by comparing unrestricted maximum Sharpe portfolios. The maximum Sharpe portfolio is of interest since the portfolio weights are determined by the estimated covariance

matrix, and has the best trade-off between risk and return. Hence we can compare covariance forecasts since the better forecast will give portfolio weights resulting in lower variance under certain assumptions.

We will also investigate if the Norwegian Pension Fund Global (SPU) can benefit from dynamic portfolio optimization. The well-known 60/40 split between stocks and bonds respectively (excluding the real estate part), is designed to capture mean reversal effect. This implies that when equity markets decline, Norges Bank Investment Management (NBIM) will have to purchase stocks. The contrarian strategy is probably well suited for SPU (Ang et al (2009)), (Fama & French (1996)). The strategy can also be derived as optimal under certain assumptions regarding return distribution and utility function. These utility functions are known as Constant Relative Risk Aversion (CRRA), it can be shown that a constant allocation to risky assets is optimal. A more in depth discussion regarding this can be found in for instance Danthine et al (2005). We do not wish to challenge the equity/fixed income rule, but to investigate the equity allocation strategy. Ang et al (2009) have shown that there are several risk premia that SPU potentially can tilt their portfolio to capture, as for example the value-growth risk and small-large risk. These strategies can still be utilized in our framework, because we only consider regional and sector indices and not specific stocks.

The Current strategy is targeting approximately: 15% Pacific, 35% America and 50% in Europe (NBIM 2012). The reason for this strategy is to maintain Norway's purchasing power with our main trading partners. This may not be optimal for the beneficiaries if a better risk return can be achieved through another strategy. To test this we construct several portfolios with different risk and return characteristics. Theory suggests that all investors should hold the tangency portfolio and then adjust their risk exposure by holding a risk free asset. This is the portfolio that gives the best risk return ratio given that there exists a risk free asset and risk is measured with standard deviation. Under classical economic assumption this is the market portfolio, because everyone is rational and the sum of everyone owns the total market. This has been shown to not always be the case, and the market portfolio is not even guaranteed to be on the efficient frontier (Gibson et al (1989)), (Fama and Macbeth (1973)). The traditional analysis only considers the two first moments of the return distribution. Research has shown that investors have preferences regarding at least the four first moments. Fama & French's portfolios with positive alphas have been shown to tilt towards recession sensitive stocks. This return characteristic is captured in the third and

fourth moment. We will therefore employ modified Conditional-Value-at-Risk to better take this into account. In this thesis modified CVaR and CVaR is used interchangeably. Behavioral finance has emerged as an important field in finance. Academics within this field have discovered that people in general assign more weight on extremely large losses with small probabilities, than small probabilities of achieving large gains see for example Benartzi and Thaler (1995). SPU's argument is that they are well suited to hold this kind of risk, because of their long investment horizon. We argue that SPU has substantial political risk and that the general opinion regarding risk tolerance may change at the worst possible time. This motivates us to apply a risk measure that takes short term risk into account. All calculations and modelling will be done through use of the statistical software R.

We try to expand the literature of applied portfolio management in the following ways: First, to explicitly focus on constructing portfolios that mimic the return or risk characteristics of the Norwegian Pension Fund Global. Second, the modified CVaR estimator with risk budgets has to our knowledge only been applied by Boudt et al (2011), but their focus was more on the general properties of this risk measure. Third, we apply both CVaR and variance as risk measure in the portfolio optimization. This is of interest because a lot of the criticism regarding the mean-variance framework is due to its lack of focus on non-normality in the return distribution which CVaR takes into account. Finally we will try to reduce trading costs using a simple technique, which will make our results more suited for real life applications.

This thesis is organized as follows: The second section derives the theoretical background for the DCC, simple moving average and the EWMA models. Section three presents the risk measures applied, followed by the covariance forecasting evaluation methodology and portfolio optimization theory in section four and five. In the sixth section we present the dataset applied followed by the results from the covariance and return evaluation in section seven. Section eight presents all portfolio optimization results, before we in section nine modify the best performing portfolio from the previous section. Finally section ten concludes based on our findings.

2. Covariance Matrix Forecasting Methods

In addition to the more complex DCC model, we will also apply simple forecasting methods such as the simple moving average and the exponential weighted moving average (EWMA) to forecast the covariance matrix.

2.1 Simple Moving Average

When predicting the covariance matrix with a simple moving average (MA) model, each return in the estimation window W_E has equal weights. For univariate series, with mean zero, the moving average variance estimate is specified as:

$$\hat{\sigma}_t^2 = \frac{1}{W_E} \sum_{k=1}^{W_E} y_{t-k}^2 \quad (1)$$

where W_E is the estimation window, y_t is the observed return on day t and σ_t^2 is the variance at day t . The estimation window is set to 100 days, as in Engle (2002).

$$Cov(Y_{t,i}, Y_{t,j}) \equiv \hat{\sigma}_{t,ij}$$

where t denotes the date and ij denotes asset i and j respectively.

In a multivariate case the covariance matrix can be forecasted the following way (Daníelsson (2011)):

$$\hat{\Sigma}_t = \frac{1}{W_E} \sum_{k=1}^{W_E} y_{t-k,ij} y'_{t-k,ij} \quad (2)$$

2.2 EWMA

Financial time series exhibit stylized facts which imply that one should assign greater weights to more recent observations. The EWMA (Risk Metrics (1996)) is based on modifying the MA so that the weights λ exponentially decline into the past.

$$\hat{\sigma}_t^2 = \frac{1-\lambda}{\lambda(1-\lambda^{W_E})} \sum_{i=1}^{W_E} \lambda^i y_{t-i}^2 \quad \text{where } \lambda \in [0,1] \quad (3)$$

where the first part of the equation ensures the weights sum to one.

The univariate EWMA can be rewritten as the weighted sum of the previous periods volatility forecast and squared returns (Daniélsson (2011)), where the sum of the weights is one:

$$\hat{\sigma}_t^2 = (1-\lambda)y_{t-1}^2 + \lambda\hat{\sigma}_{t-1}^2 \quad (4)$$

The multivariate form of the model is almost the same:

$$\Sigma_t = \lambda\Sigma_{t-1} + (1-\lambda)y_{t-1}y'_{t-1} \quad (5)$$

where Σ is the covariance matrix, and y_{t-1} is a return vector lagged one period.

With individual elements given by:

$$\sigma_{t,ij} = \lambda\sigma_{t-1,ij} + (1-\lambda)y_{t-1,i}y'_{t-1,j}, \quad i, j = 1, \dots, K \quad (6)$$

2.3 Univariate GARCH Model

Returns of financial assets tend to be correlated, and the volatility of assets tends to cluster. Hence, modelling volatility, conditional on previous returns should give a better estimate of tomorrow's volatility than an unconditional volatility forecast. Understanding how the univariate generalized autoregressive conditional heteroskedasticity (GARCH) model works is important when working with the dynamic conditional correlation (DCC) model, since it basically is a nonlinear combination of univariate GARCH models.

The error process is given by: $\varepsilon_t = v_t\sqrt{h_t}$

$$\sigma_v^2 = 1$$

$$h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i} \quad (7)$$

$\{v_t\}$ is a strict white noise process causing the conditional and unconditional means of ε_t are equal to zero.

$$E\varepsilon_t = Ev_t(h_t)^{0.5} = 0 \quad (8)$$

The conditional variance is given by:

$$E_{t-1}\varepsilon_t^2 = h_t \quad (9)$$

Thus the conditional variance of ε_t^2 is the ARMA process given by h_t .

2.4 DCC-GARCH

2.4.1 The DCC-GARCH Model

The DCC model introduced by Engle (2001) assumes that the time series has zero mean and no autocorrelation. If this is not the case, the data is prewhitened by an ARMA-model. In our study we have used the residuals from a fitted ARMA-model and the covariance matrix is specified as:

Returns: $r_t | \mathbb{R}_{t-1} \sim N(0, H_t)$

Covariance matrix: $H_t \equiv D_t R_t D_t$

where D_t is the $k \times k$ diagonal matrix of time varying standard deviations from univariate GARCH models with $\sqrt{h_{it}}$ on the i^{th} diagonal, and R_t is the time varying correlation matrix.

The elements of D_t can be written as univariate GARCH models:

$$h_{it} = \omega_i + \sum_{p=1}^{P_i} \alpha_{ip} r_{it-p}^2 + \sum_{q=1}^{Q_i} \beta_{iq} h_{it-q} \quad (10)$$

For $i = 1, 2, 3, \dots, k$ with the GARCH restrictions such as non-negativity of variances and stationarity ($\sum_{p=1}^{P_i} \alpha_{ip} + \sum_{q=1}^{Q_i} \beta_{iq} < 1$) being imposed. The lag lengths for P and Q do not need to be the same, and the univariate GARCH models can include any GARCH process with normally distributed errors which satisfies the stationarity and non-negativity

constraints, we have applied the traditional GARCH(1,1). The dynamic correlation structure is formulated as:

$$Q_t = \left(1 - \sum_{m=1}^M \alpha_m - \sum_{n=1}^N \beta_n\right) \bar{Q} + \sum_{m=1}^M \alpha_m (\epsilon_{t-m} \epsilon'_{t-m}) + \sum_{n=1}^N \beta_n Q_{t-n} \quad (11)$$

$$R_t = Q_t^{*-1} Q_t Q_t^{*-1} \quad (12)$$

where \bar{Q} is the unconditional covariance of the standardized residuals from the first stage estimation. And R_t is the conditional correlation matrix where a typical element is on the

form: $\rho_{ij,t} = \frac{q_{ij,t}}{\sqrt{q_{ii}q_{jj}}}$

$$Q_t^* = \begin{bmatrix} \sqrt{q_{11,t}} & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{q_{22,t}} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sqrt{q_{kk,t}} \end{bmatrix}$$

where Q_t^* is a diagonal matrix consisting of the square root of the diagonal elements of Q_t .

2.4.2 Estimation of the DCC(1,1) Model

The DCC model can be estimated in two stages, where in the first stage univariate GARCH models are estimated for each series of residuals and in the second stage the transformed residuals from the first stage are used to estimate the dynamic correlation parameters.

The log-likelihood of this estimator can be written:

$$L = -\frac{1}{2} \sum_{t=1}^T (k \log(2\pi) + \log(|H_t|) + r_t' H_t^{-1} r_t) \quad (13)$$

$$= -\frac{1}{2} \sum_{t=1}^T (k \log(2\pi) + \log(|D_t R_t D_t|) + r_t' D_t^{-1} R_t^{-1} D_t^{-1} r_t) \quad (14)$$

$$= -\frac{1}{2} \sum_{t=1}^T (k \log(2\pi) + 2 \log(|D_t|) + \log(|R_t|) + \epsilon_t' R_t^{-1} \epsilon_t) \quad (15)$$

where $\epsilon_t \sim N(0, R_t)$ are the residuals standardized by their conditional standard deviation.

Step 1:

The first stage estimation involves replacing R_t with I_k , an identity matrix of size k . The model parameters, θ , is written in two groups $(\varphi_1, \varphi_2, \dots, \varphi_k, \psi) = (\varphi, \psi)$, where the elements of φ_i correspond to the parameters of the univariate GARCH model for the i^{th} asset series, $\varphi_i = (\omega, \alpha_{1i}, \dots, \alpha_{p_i i}, \beta_{1,i}, \dots, \beta_{q_i i})$. The first step quasi likelihood function is then specified as:

$$\begin{aligned} QL_1(\varphi|r_t) &= -\frac{1}{2} \sum_{t=1}^T (k \log(2\pi) + \log(|I_k|) + 2 \log(|D_t|) + r_t' D_t^{-1} I_k D_t^{-1} r_t) \\ &= -\frac{1}{2} \sum_{t=1}^T (k \log(2\pi) + 2 \log(|D_t|) + r_t' D_t^{-2} r_t) \\ &= -\frac{1}{2} \sum_{t=1}^T (k \log(2\pi) + \sum_{n=1}^k (\log(h_{it}) + \frac{r_{it}^2}{h_{it}})) \\ &= -\frac{1}{2} \sum_{n=1}^k (T \log(2\pi) + \sum_{t=1}^T (\log(h_{it}) + \frac{r_{it}^2}{h_{it}})) \end{aligned} \quad (16)$$

This is the sum of the log likelihoods of the individual GARCH models for each asset.

In the second step the correctly specified log-likelihood function is used to estimate the parameters, given the parameters estimates from the first stage likelihood.

Step 2:

$$\begin{aligned}
QL_2(\psi|\hat{\varphi}, r_t) &= -\frac{1}{2} \sum_{t=1}^T (k \log(2\pi) + 2 \log(|D_t|) + \log(|R_t|) + r_t' D_t^{-1} R_t^{-1} D_t^{-1} r_t) \\
&= -\frac{1}{2} \sum_{t=1}^T (k \log(2\pi) + 2 \log(|D_t|) + \log(|R_t|) + \epsilon_t' R_t^{-1} \epsilon_t) \tag{17}
\end{aligned}$$

Since the two first terms are constants and we are conditioning on $\hat{\varphi}$, only $\log(|R_t|) + \epsilon_t' R_t^{-1} \epsilon_t$ will influence the parameter selection. So when estimating the DCC parameters the log likelihood function can be written:

$$QL_2(\psi|\hat{\varphi}, r_t) = -\frac{1}{2} \sum_{t=1}^T (\log(|R_t|) + \epsilon_t' R_t^{-1} \epsilon_t) \tag{18}$$

2.4.3 Forecasting the Covariance Matrix

Maybe the most important application of the DCC model is to forecast the covariance matrix. Engle (2001) shows how the DCC model can be applied to do multi-steps-ahead forecasts, but since we will optimize the portfolio every period, we only consider the one-step-ahead forecast. This can be computed in the following way:

$$Q_{t+1} = (1 - \alpha - \beta)\bar{Q} + \alpha(\epsilon_t \epsilon_t') + \beta Q_t \tag{19}$$

3. Measuring Risk

3.1 Variance

Variance is a measure that captures volatility, the logic behind this measure is that for series with higher variance, investors are more uncertain concerning tomorrow's return, and thus investors wish to be compensated accordingly. One of the cornerstones in finance is that diversification gains can be obtained due to different price impacts to economic shocks. The portfolio's variance is therefore not equal to the sum of individual asset's variance. We use the conventional portfolio variance definition:

$$\sigma_{p,t}^2 = \text{var}(\tilde{R}_{p,t}) = (w' \Sigma w)_t \quad (20)$$

where Σ is the covariance matrix, and w is a weight vector.

3.2 Value at Risk

We follow the notation used in Danielsson (2011), and start with defining log return.

$$Y_t = \log(P_t) - \log(P_{t-1})$$

where P_t is the price

$$p = \Pr(P_{t-1}(e^{Y_t} - 1) \leq -VaR(p))$$

$$p = \Pr(P_{t-1}(e^{Y_t} - 1) \leq -VaR(p))$$

$$p = \Pr(P_{t-1}e^{Y_t} \leq -VaR(p) + P_{t-1})$$

$$p = \Pr\left(e^{Y_t} \leq -\frac{VaR(p)}{P_{t-1}} + 1\right)$$

$$p = \Pr\left(Y_t \leq \log\left(-\frac{VaR(p)}{P_{t-1}} + 1\right)\right)$$

$$p = \Pr\left(\frac{Y_t}{\sigma} \leq \log\left(-\frac{VaR(p)}{P_{t-1}} + 1\right)\left(\frac{1}{\sigma}\right)\right) \quad (21)$$

since $-\frac{VaR(p)}{P_{t-1}} \leq 1$ the function is defined. The distribution of standardized returns (Y_t/σ) can be denoted $F_y(\cdot)$ and the inverse distribution by $F_y^{-1}(p)$ we have:

$$\log\left(-\frac{VaR(p)}{P_{t-1}} + 1\right) = F^{-1}(p) * \sigma$$

$$VaR(p) = (\exp((F^{-1}(p) * \sigma)) - 1) * P_{t-1}$$

For small $F^{-1}(p) * \sigma$ the VaR is approximately given as:

$$VaR \approx -\sigma * F^{-1}(p)P_{t-1} \quad (22)$$

For a more thorough derivation and definition of higher moments, please see appendix D.

3.3 Conditional Value at Risk

Conditional value at risk (CVaR) is unlike value at risk (VaR) a coherent risk measure and is also a convex function of the portfolio weights. (Rockafellar and Uryasev, (2000); Artzner et al. (1999). We chose to define CVaR in percentage returns since our goal is to compare portfolios based on assets returns.

The Definition of CVaR is “Expected loss conditional on VaR being violated.” (Financial risk forecasting):

We can define the expectation the following way:

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

$f(\cdot)$ is defined as the probability density function, and $f_Q(\cdot)$ has support on the interval

$$[-\infty, -VaR(p)]$$

$$CVaR = -E[Q|Q \leq -VaR(p)]$$

where Q is defined as the expected profit/loss.

$$= \int_{-\infty}^{-VaR(p)} x f_{(Q|Q < -VaR(p))}(x) dx$$

where $f_{(Q|Q < -var(p))}$ can be found the following way:

$$\begin{aligned} F_{Q|Q < -var(p)}(x) &= P(Q \leq x | Q < -VaR(p)) \\ &= \frac{P(Q \leq x)}{P(Q < VaR(p))} = \frac{F_Q(x)}{p} \end{aligned}$$

This implies that the derivative (the pdf) is:

$$f_{(Q|Q < -VaR(p))} = \frac{f_Q(x)}{p} \tag{23}$$

3.3.1 CVaR under Normality:

$$CVaR = - \int_{-\infty}^{-VaR(p)} x f_Q(x) dx$$

$$CVaR = \frac{1}{p} \int_{-\infty}^{-VaR(p)} x * \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) \exp \left[-\frac{1}{2} * \frac{x^2}{\sigma^2} \right] dx$$

$$CVaR = \frac{1}{p} \left[- \left(\frac{\sigma^2}{\sqrt{2\pi\sigma^2}} \right) \exp \left[-\frac{1}{2} * \frac{x^2}{\sigma^2} \right] \right]_{-\infty}^{-VaR(p)}$$

The bracket only needs to be evaluated at the boundaries, since the lower bound is approximately zero, the standard normal density function is: $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp[-\frac{x^2}{2}]$ which gives us:

$$CVaR = - \frac{\sigma^2 \phi(-VaR(p))}{p} \quad (24)$$

Financial returns are seldom normally distributed which implies that the risk contribution to CVaR should be calculated in a way that takes this into account. There are basically two ways to compute the non-normal risk contribution to CVaR. First one can find the expected CVaR contribution by using historical or simulated data. The downside of the historical data approach is that it demands a very large sample. For example, when computing the 1% CVaR one should have at least a sample size of 1000, the 1% CVaR is then calculated based on the 10 smallest observations, Daníelsson (2011). A more elegant approach to calculate the risk contributions is to derive an analytical formula which takes into account the non-normal distribution of the returns. In this thesis we will apply the modified CVaR estimator of Boudt et al. (2008).

3.4 The Modified CVaR Estimator

The modified CVaR estimator is based on the Cornish-Fisher expansion and is a function of the co-moments of the underlying asset returns, the estimator has been shown to give accurate estimates of the CVaR contributions. This allows for a more realistic approximation of the true distribution. The Cornish Fisher expansion will be identical to the normal distributions in the case where skewness is 0 and kurtosis equals 3. We have used historical estimation of the third and fourth co-moments, while the forecasted covariance is used as the second moment. Throughout this thesis we will set the loss probability α to 5% as is common in practice. Especially higher moments are very sensitive to extreme observations, and we therefore “cleaned” the dataset to get more robust estimates. For details, see Appendix A.

3.4.1 The Cornish-Fisher Expansion

The Cornish fisher expansion can be used to derive approximates to quantiles, utilizing higher moments of the actual distribution (Cornish and Fisher (1937)).

$$z_{cf} = z_{\alpha} + \frac{(z_{\alpha}^2 - 1)S}{6} + \frac{(z_{\alpha}^3 - 3z_{\alpha})K}{24} - \frac{(2z_{\alpha}^3 - 5z_{\alpha})S^2}{36} \quad (25)$$

z_{α} is the α percentile of the standard normal distribution. Using the Cornish Fisher expansion, CVaR can be approximated the following way:

$$MCVaR(p) = -\sigma \left(M_1 + \frac{(M_2 - 1)S}{6} + \frac{(M_3 - 3M_1)K}{24} - \frac{(2M_3 - 5M_1)S^2}{36} \right) \quad (26)$$

where $M_i = \frac{1}{p} \int_{-\infty}^{-VaR(p)} x^i f(x)$ and $f(\cdot)$ is the standard normal probability density function (Cao et. al. (2009)). This approximation fits best when kurtosis and skewness only deviates

moderately from the normal distribution. This approach also has wrong tail behaviour i.e. when α goes to 0, CVaR tends to zero (Boutd et al (2008)).

To avoid results where CVaR is smaller than mVaR the following definition of CVaR is used:

$$mCVaR^* = -\sigma \min\{CVaR, VaR\} \quad (27)$$

4. Methodology for Covariance and Return Evaluation

4.1 Covariance Forecast Evaluation

In order to find out which of the methods (MA, EWMA and DCC) produced the best forecast. We employed the test developed by Engle and Colacito (2006) (EM-test). They proved that the covariance is smallest for the best specified covariance forecast.

The test adopts the classical portfolio, that an investor minimizes the variance for a required rate of return.

$$\min_{w_t} w_t' \hat{\Sigma}_t w_t \quad (28)$$

$$\text{Subject to } w_t' \mu = \mu_0$$

The solution to this problem is given by:

$$w_t = \frac{\hat{\Sigma}_t^{-1} \mu}{\mu' \hat{\Sigma}_t^{-1} \mu} \mu_0 \quad (29)$$

where μ is a vector of excess return over the risk free rate and μ_0 is the required return. Note that we do not require the weight's to sum to one, because one minus the sum of the weights is allocated to the risk free asset. This is the classical portfolio optimization where part of the portfolio is invested in the tangency portfolio. If weights are rescaled to one, we would find an unconstrained tangency portfolio. In order to isolate the effect from the covariance forecast, the expected return is constant and equal to its historical average. The return target is set equal to the average excess return and as a proxy for risk free rate we have used the three months Treasury bill yield (0.09% 13 April 2012).

Engle and Colacito (2006) showed that if we know the true covariance forecast Ω_t any weights constructed from another covariance forecast will produce higher or equal standard deviation standardized by required return.

$$\frac{\sqrt{w_t' \Omega_t w_t}}{\mu_0} \geq \frac{\sqrt{(w_t^*)' \Omega_t w_t^*}}{\mu_0} \quad (30)$$

w_t^* is optimal weights obtained from the minimization procedure, given that we know the true covariance. This can be reduced to:

$$\sigma_t \geq \sigma_t^* \quad \forall \Sigma_t \neq \Omega_t \quad (31)$$

where σ_t^* is the standard deviation obtained from the true forecast. Engle and Colacito (2006) expand this result to comparing two competing covariance matrix estimates, and prove that the covariance matrix that obtains the lowest variance is closest to the true covariance.

The test computes portfolio return π_t^j by:

$$\pi_t^j = (w_t^j)' (r_t - \bar{r})$$

where $w_t^{j,k} = \frac{(\bar{\Sigma}_t^j)^{-1} (\mu)'}{(\mu)' (\bar{\Sigma}_t^j)^{-1} \mu} \mu_0$ denotes the weights obtained from covariance forecast j. r_t is the return that, \bar{r} is the mean return.

$$u_t = (\pi_t^a)^2 - (\pi_t^b)^2, \quad t = 1, \dots, T$$

Given that the mean is zero the square of π can be viewed as the portfolio variance. The null hypothesis is that u_t^k is null for all k. This is a Diebold and Mariano (1995) and assess if u_t is significantly different from zero. The test is to regress u_t^k on a constant using generalized method of moments with a Newey West covariance matrix. The reason for this is to correct for possible problems concerning heteroskedasticity, autocorrelation and non-normality.

$$u_t = \beta_u \iota + \epsilon_t$$

where ι is a $k \times 1$ vector, and β_u is a scalar.

4.2 Expected Return

Our main focus in this thesis is on risk prediction, but the literature suggests there is a significant momentum effect (Scowcroft & Sefton (2005)). This is mainly found in sector indices, and suggests that the use of an ARMA model may be appropriate to predict expected return. The majority of the momentum effect in their portfolio was realized by long positions, so we should still be able to capture part of the momentum effect with a no shorting restriction. This can of course produce negative expected returns, which are counter intuitive, but the optimization will still seek to find the lowest risk given a return target. Given our myopic optimization; this is probably a better approach than a mean reversal strategy found over longer horizons (Cochrane (1999)).

We employed the DM test in order to test if we can predict return better with an ARMA model than with a 12 month moving average model, which was shown to be the best momentum predictor.

$$e_{k,t}^2 = (r_{k,t} - \hat{r}_{k,t})^2 \quad (1.1)$$

where $e_{k,t}^2$ is the mean squared error, of forecast k . We then construct:

$$d_t = e_{1,t}^2 - e_{2,t}^2 \quad (1.2)$$

The null hypothesis is that d_t is zero. We regress d_t on a constant, and use heteroskedastic, autocorrelation consistent standard errors in order to test this.

$$d_t = \beta \iota + \epsilon_t$$

where ι is a $k \times 1$ vector, and β is a scalar.

5. Portfolio Optimization Theory

We will construct the following portfolios: Minimum variance, maximum Sharpe, risk budgets equal to SPU's strategic weights, risk budgets equal to sector market capitalization weights and minimum CVaR/variance with a return target equal to SPU's mean expected return.

Extreme negative weights may occur in efficient portfolios, it would then appear that imposing a non-negativity portfolio weight constraint would lead to a loss in efficiency. However empirical findings in this area suggest that imposing these constraints on portfolio weights improve the efficiency. See Frost and Savarino (1988) for an excellent discussion.

The unconstrained optimization is often shown to produce corner solutions, where an extremely large part of the portfolio is allocated to a single asset. A common technique called shrinkage that is often applied, reduces the impact from extreme estimates. Jagannathan and Ma (2003) demonstrated that no shorting constrained portfolios work almost as well. We will therefore implement the realistic restriction of no shorting. Expected return is assumed to follow an ARMA process, and we further assume that there is no risk free asset, except for the max Sharpe portfolio. The last assumption has the implication that no tangency portfolio can be found, and thus the entire efficient frontier is optimal, and is only depending on the agent's risk aversion. We do not wish to place any explicit assumptions regarding utility function or risk aversion, but instead assume that The Norwegian Pension Fund Global's strategic weights reflect their risk preference.

5.1 Minimum Variance

Minimum variance has recently prompted great interest both from academic researchers and market practitioners, as the construction does not rely on expected returns and is therefore assumed to be more robust. (Maillard et. Al. (2008)). Merton (1980) showed that small changes in expected returns, can lead to significant variations in the composition of the portfolio. This is of great interest for our study because large turnover is an unfeasible option for a fund which owns approximately 1% of the global stock market (Reuters 2009).

The standard criticism regarding minimum variance is that it tends to be biased toward value and small-firm effect (NBIM (2012)). This is not a feasible outcome because no individual stocks are considered.

The Global mean variance portfolio (GMVP) is computed as the solution to:

$$\min_{w_t} w_t' \hat{\Sigma}_t w_t$$

$$\text{Subject to } w_t' \iota = 1$$

$$0 < w_i < 1$$

where ι is a summation vector.

5.2 Maximum Sharpe

The classical model assumes that all investors would want to hold the maximum Sharpe portfolio (tangency portfolio) (Sharpe (1964)). The optimization for max Sharpe is identical to that employed in EM test, but with a no shorting constraint. The portfolio is then divided by the sum of weights, to ensure that it's fully invested in equity. Expected returns are allowed to change with the ARMA forecasts.

5.3 Minimizing variance and CVaR with a Return Target

The rationale for this portfolio is to achieve the optimal asset allocation given the same return target as the Norwegian Pension Fund Global. The return target is thus designed as the expected return given the strategic weights, i.e.

$$Target = (w_t^{strategic})' r_t \tag{32}$$

where w_t is a weight vector, and r_t is the forecasted return from the ARMA process. All other constraints are equal to the minimum variance portfolio. For both return target portfolios we applied an r-code¹ for constrained portfolio optimization.

5.4 Risk Budgets with CVaR and Variance

In this part we have constrained the risk budgets (RB) to be equal to the Norwegian Pension Fund Global strategic weights or market capitalization weights for sectors. This ensures that these portfolios have a risk exposure which is equal to the strategic weights. These portfolios also have the advantage of not depending on expected return. The benefit of this approach is that minimum variance often produces heavy weighting to some assets (Maillard et. Al. (2008)). This can lead to overexposure to political (idiosyncratic) risk in certain regions or to certain industries. The RB ensures that the portfolio is well diversified across investment opportunities. Qian (2006) showed that the decomposition of risk can be a significant predictor of each asset (ex-post) losses.

The optimization procedure is identical to minimum variance, with constraints on risk contribution and the derivation of the risk contribution for variances is straight forward, and can be done the following way because the covariance matrix is a symmetric matrix.

$$\nabla_w = \frac{\partial}{\partial w} (w' \Sigma w) = 2 \Sigma w$$

The derivative of the standard deviation (σ) is then:

$$\frac{\partial \sigma}{\partial w} = 0.5 * \sigma^{-0.5} (2 \Sigma w) = \frac{\Sigma w}{\sqrt{\sigma}} \quad (33)$$

The marginal percentage contribution from each asset is therefore:

$$\frac{\partial \sigma / \partial w_i}{\sigma} * w_i \quad (34)$$

The derivative of CVaR is more tedious, and we therefore refer interested readers to Boudt et al (2008) Appendix C.

¹<https://r-forge.r-project.org/scm/viewvc.php/pkg/optimizer/R/optimize.portfolio.R?view=markup&root=returnanalytics&pathrev=1433>

The marginal percentage contribution can be written the following way:

$$\frac{(\partial CVaR / \partial w_i)}{CVaR} * w_i \quad (35)$$

The risk budget portfolio usually outperforms the market index, due to low-volatility anomaly and business cycle component (NBIM (2012)).

We applied an optimization method called Differential Evolution for both the CVaR and Variance risk contribution portfolios. This is because CVaR and variance with risk budgets is not necessarily a convex function of the portfolio weights, and may also be non-differentiable. The DE algorithm is derivative free global optimizer, which allows for risk restrictions (Boudt et al (2009)). For details please see Appendix E.

5.5 Regional and Sector Rebalancing Strategies

As a benchmark for the more complex portfolios, we constructed portfolios which were rebalanced by a trigger strategy. For the regional investment universe this was done around SPU's current strategic regional weights (50% Europe, 35% North America, 15% Pacific) (NBIM (2011)). The trigger was set to 3 percentage points, meaning that every time a regional weight exceeds the strategic regional weights by this much, it will be rebalanced back to the strategic weights. For an overview of rebalancing strategies, see NBIM (2012). The trigger for the sector portfolio where also set to 3 percentage points.

The Norwegian Global Pension Fund (SPU) is currently rebalanced around fixed regional weights, how this is exactly done is currently not public available information. From 1998-2001 the fund where rebalanced back to the original regional weights every quarter, thus by a calendar-based rule (Norges Bank (2012)). Since 2001 the rebalancing regime has consisted of two elements, partly and full rebalancing. Partly rebalancing has followed the monthly supply of new capital and the regional weights have been adjusted in direction of the original regional weights. The full rebalancing has been a decision based on the current deviations from the regional weights. In a letter dated 26 of January 2012 the Norwegian Bank suggests that the rebalancing of the fund should be done based on the asset allocation between stocks and bonds with a threshold of three percent. Specifically this means that if

the equity value of the fund exceeds 63% or below 57%, the equity weight will be brought back to 60% in the end of the nearest quarter. The practice of partly rebalancing will be discontinued.

6. Description of the Dataset

We will calculate all portfolios for two different investment universes, one where we divide the world into regions, and one where we separate the investment universe into different industry sectors. The data applied are all daily total return indices which are split into the in-sample period; 15.06.1995-29.12.1999 for regions and 04.03.1996-31.12.2012 for sectors and the out-of-sample period 04.01.2000-12.04.2012 for both. The in sample estimation periods both consists of 1000 observation. Those observations where one or more indices were not updated were removed. All data used in our analysis is retrieved through Datastream and the indices applied are delivered by MSCI, for a thorough explanation of how the indices are put together, please visit MSCIs websites.

6.1 Regions

We chose to divide the world into seven different regions; Europe, Japan, Africa, North America, Latin America, India and Pacific.ex Japan. The rationale for this division is both good possibilities of diversification and that we seek to mimic the SPU strategy of regional rebalancing. See figure 2 for SPU's strategic weights.

The regions exhibit quite different characteristics when measured by mean and risk, where risk is defined as both CVaR and standard deviations. See figure 1 and table 1. All regional returns are almost symmetrical distributed but they generally exhibit large positive Kurtosis, indicating that the possibilities of extreme negative and positive outcomes exceeding those of the normal distribution. Table 2 shows that the correlation between the regions are in the range between 0,2 and 0,64. We also notice that mean return in this period for the developed world (i.e. Nort America, Europe and Japan) is negative over this period, demonstrating that a strategy were SPU is heavily weighted in these regions may not always be optimal. South Africa is used as a proxy for Africa, and show an high Sharpe ratio.

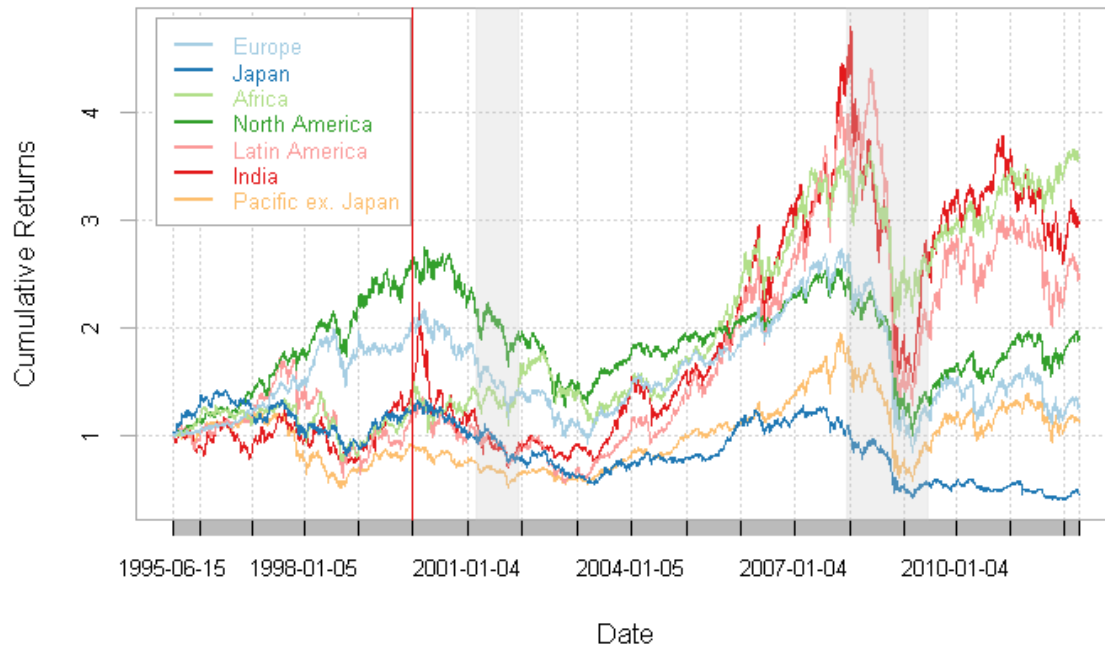


Figure 1: Cumulative returns for all regions for the entire period; 1995.01.04 – 2012.04.12. The red line marks the transition from in to out-of-sample. The shaded area indicates a period of economic contraction (NBER (2012)).

	Europe	Japan	Africa	North America	Latin America	India	Pacific ex. Japan
Annual Mean	-1.04 %	-6.58 %	11.80 %	-0.41 %	12.14 %	11.24 %	5.06 %
Standard Deviation	26.35 %	24.48 %	23.18 %	22.86 %	32.89 %	29.79 %	25.01 %
Annual Sharpe	-0.07	-0.30	0.47	-0.06	0.34	0.35	0.17
Skewness	0.00	-0.26	-0.06	-0.21	-0.12	-0.03	-0.06
Kurtosis	8.09	5.98	4.14	7.81	10.58	6.18	11.39
CVaR	-3.50 %	-4.28 %	-3.51 %	-3.75 %	-4.14 %	-4.35 %	-2.65 %

Table 1: Descriptive statistics for all regional returns for the out of sample period; 04.01.2000 – 04.01.2012.

	Europe	Japan	Africa	North America	Latin America	India	Pacific ex. Japan
Europe	1.00						
Japan	0.38	1.00					
Africa	0.57	0.37	1.00				
North America	0.59	0.22	0.36	1.00			
Latin America	0.64	0.31	0.48	0.66	1.00		
India	0.33	0.32	0.34	0.20	0.30	1.00	
Pacific ex. Japan	0.60	0.61	0.52	0.34	0.48	0.43	1.00

Table 2: Unconditional correlations for all regional returns for the entire sample period; 04.01.1995 – 04.01.2012.

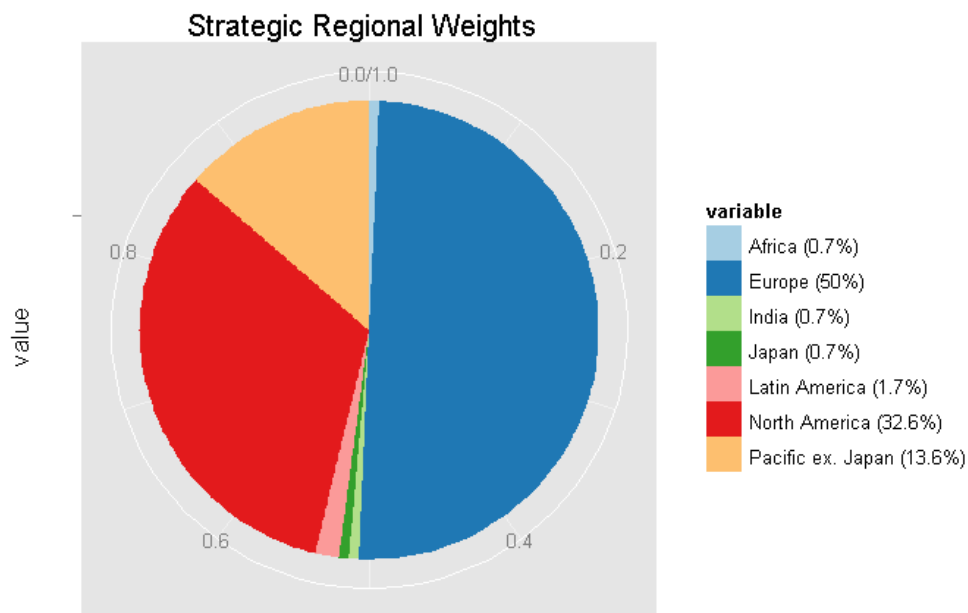


Figure 2: The figure shows SPU's strategic regional weights (50% in Europe, 35% in America and Africa and a total of 15% in India, Japan and Pacific ex Japan).

6.2 Sectors

We further divided the global investment universe into ten different sectors; Consumer, Discretionary, Consumer Staples, Energy, Financials, Health Care, Industrials, Information Technology, Materials, Telecommunication Services and Others. See figure 4 for actual market capitalization weights. The reason for choosing this subdivision was to test if this could deliver better diversification possibilities, but from table 4 there seems obvious that the sectors exhibit higher correlation than the regional subdivision (between 0,44 -0,9 versus 0,2-0,64 for regions).

The different sectors also display quite different characteristics when measured by mean and risk, where risk is defined as both CVaR and standard deviations. See figure 3 and table 3. There seems to some differences in skewness, where negative skewness increases the probability for extreme negative outcomes. The indices have high kurtosis (peaked distribution). This motivates a risk measure that is able to capture non normality in the return distribution.

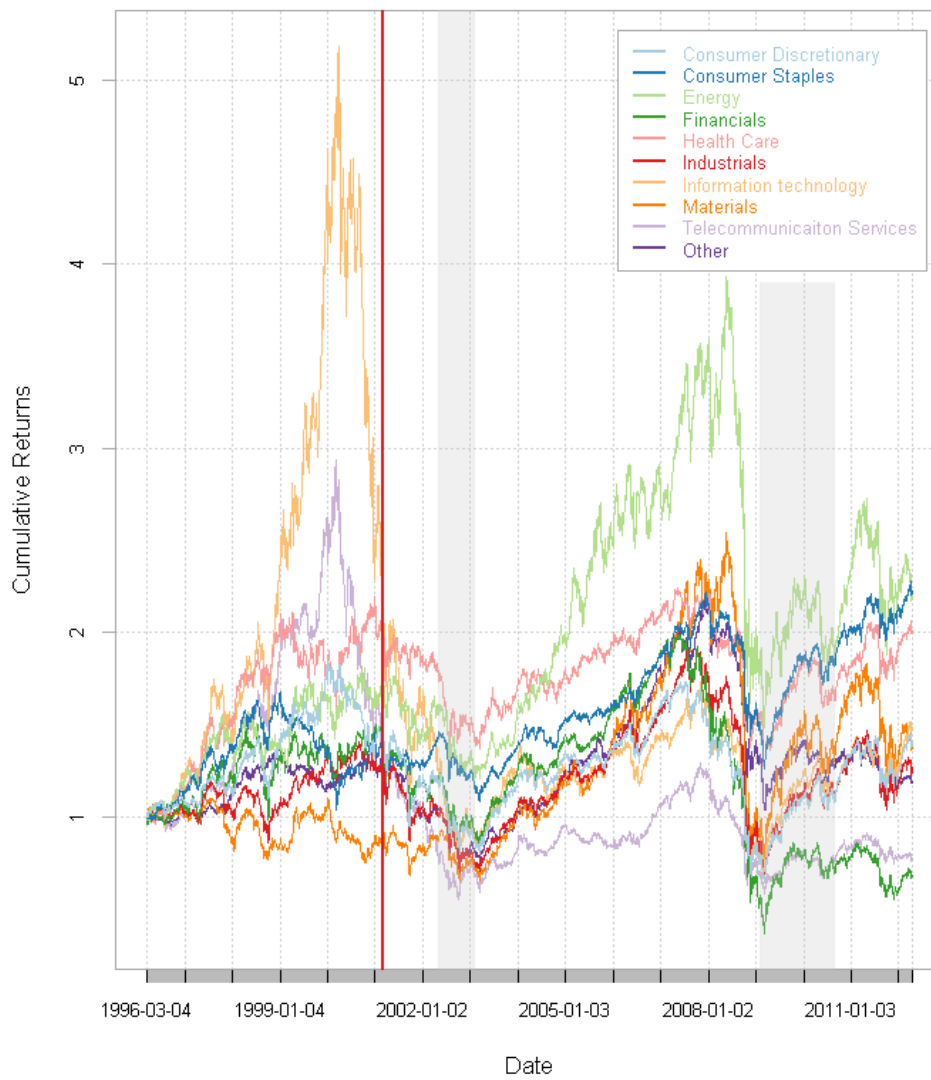


Figure 3: Cumulative returns for all sectors for the entire period; 1995.01.04 – 2012.04.12. The red line marks the transition from in to out-of-sample.

	Consumer Discretionary	Consumer Staples	Energy	Financials	Health Care
Annual Mean	-0.44 %	4.95 %	5.69 %		-2.96 %
Standard Deviation	19.01 %	13.53 %	25.39 %		24.00 %
Annual Sharpe	-0.07	0.30	0.19		-0.16
Skewness	0.12	-0.27	-0.55		-0.10
Kurtosis	7.33	8.61	8.99		8.98
CVaR	-2.33 %	-2.27 %	-5.18 %		-3.38 %

	Industrials	Information technology	Materials	Telecommunication Services	Other
Annual Mean	1.40 %	-5.47 %	5.08 %	-7.89 %	1.43 %
Standard Deviation	19.54 %	26.95 %	23.63 %	20.79 %	16.02 %
Annual Sharpe	0.03	-0.24	0.18	-0.42	0.03
Skewness	-0.37	0.11	-0.47	-0.02	-0.14
Kurtosis	4.97	3.98	7.54	4.12	12.39
CVaR	-3.55 %	-3.73 %	-4.65 %	-3.15 %	-1.72 %

Table 3: Descriptive statistics for all sector returns for the out of sample period sample period; 04.01.2000 – 04.01.2012.

	Consumer Discretionary	Consumer Staples	Energy	Financials	Health Care	Industrials	Info- tech	Materials	Telecom Services	Other
Consumer Discretionary	1									
Consumer Staples	0.67	1								
Energy	0.59	0.61	1							
Financials	0.83	0.7	0.64	1						
Health Care	0.64	0.77	0.56	0.66	1					
Industrials	0.9	0.71	0.67	0.87	0.67	1				
Information technology	0.74	0.45	0.43	0.62	0.5	0.71	1			
Materials	0.75	0.64	0.72	0.77	0.55	0.84	0.51	1		
Telecommunicaiton Services	0.72	0.59	0.53	0.7	0.57	0.7	0.63	0.6	1	
Other	0.65	0.71	0.69	0.69	0.64	0.7	0.44	0.69	0.64	1

Table 4: Unconditional correlations for all regional returns for the entire sample period; 04.03.1996 – 04.01.2012.



Figure 4: The figure shows the global market capitalization of different sectors.

7. Results from the Covariance and Return Evaluation

7.1 Fitting Models

As input in the DCC model one has to obtain the residuals from a fitted time series model, also known as prewhitening the time series. This is important because the model assumes no linear autocorrelation. We will apply the Autoregressive moving average (ARMA) model, introduced by Box and Jenkins (1976) to obtain residuals used in the covariance forecast, and to predict expected returns. The model is specified as (Walter Enders (2010)):

If one combines a moving average process with a linear difference equation one obtains the ARMA model.

$$y_t = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i} + \sum_{i=0}^q \beta_i \varepsilon_{t-i}$$

where the first part is the autoregressive terms and the second is the moving average terms. Appropriate lag length for the AR and MA part were determined by BIC information criteria which has been shown to be more asymptotically correct than AIC.

After the ARMA models were fitted based on the BIC information we investigated the autocorrelation plots, and found there to be some significant autocorrelation in the residuals. However there is a trade-off between parsimony (i.e. robust forecasts) and models with more parameters (produces residuals closer to white noise) (Walter Enders (2010)) where the first is important for the expected return prediction. See appendix C for diagnostic plots of residuals.

Secondly we fitted univariate GARCH(1,1) models in order to capture the volatility, which is used in the first step of the DCC model. Standardized returns from these models should not exhibit any kind of autocorrelation. The white noise process is not directly observable so we used the estimated counterpart $\hat{v}_t = \frac{\hat{\varepsilon}}{\sqrt{\hat{h}_t}}$. The ACF plots are found in appendix C and confirm that this is in fact the case. This is essential to ensure that we utilize all information in past returns, \hat{v}_t should also have mean 0 and variance 1. These statistics are reported in the table below.

	Europe	Japan	Africa	North America	Latin America	India	Pacific ex. Japan
Mean	0.00	-0.01	-0.01	0.00	-0.01	0.00	-0.01
Variance	0.99	1.00	1.00	0.99	1.00	0.99	1.00

Table 5: Mean and variance for standardized residuals.

We notice that all means and variance are close to 0 and 1 respectively. v_t and its estimated counterpart \hat{v}_t is assumed to be normally distributed.

	Consumer Discretionary	Consumer Staples	Energy	Financials	Health Care
Mean	-0.01	-0.01	-0.01	-0.01	-0.01
Variance	1.00	0.99	0.99	1.00	0.99

	Industrials	Information technology	Materials	Telecom. Services	Other
Mean	-0.01	-0.01	-0.01	-0.01	-0.01
Variance	1.00	0.99	0.99	1.00	0.99

Table 6: Mean and Variance for standardized residuals.

The same seems to be true for sectors.

7.2 Expected Return

We clearly see a much higher tendency for a significant first lag autocorrelation in sectors than when the world is divided into regions. This is consistent with the momentum theory. And the process does not seem to be a complete random walk.

Results Regions:

	Estimate	t value
Europe	9.20E-07	0.722
Japan	1.26E-07	0.114
Africa	-1.22E-06	-1.377
North America	1.24E-06	0.568
Latin America	-5.23E-06	-2.462
India	-3.13E-06	-1.755
Pacific ex. Japan	-2.33E-06	-1.612

Table 7: Results from the DM test, a negative and significant estimate indicates that the ARMA model gives a better forecast of the return than the moving average model. Significant Coefficients are marked with two stars at the 5% significance level, and, and one at the 10%.

Results Sector:

	Estimate	t value
Consumer Discretionary	-4.36E-06	-4.074**
Consumer Staples	-6.54E-07	-1.708*
Energy	-6.05E-07	-0.243
Financials	-1.80E-06	-0.738
Health Care	-1.54E-06	-2.152**
Industrials	-1.71E-06	-1.605
Information technology	-2.95E-07	-0.203
Materials	-7.57E-06	-2.420**
Telecommunication Services	-4.21E-07	-0.381
Other	-6.47E-07	-1.408

Table 8: Results from the DM test, a negative and significant estimate indicates that the ARMA model gives a better forecast of the return than the moving average model. Significant Coefficients are marked with two stars at the 5% significance level, and, and one at the 10%.

As we can see most of the betas are negative, and for the sector subdivision, four of them are statistically significant, indicating that we can capture momentum better with an ARMA forecast, than with a 12 month moving average model.

We will thus use ARMA forecast as expected return throughout the rest of this thesis.

7.3 Evaluating the Performance of Different Covariance Forecasting Methods

The table reports coefficients and t-statistics, and indicates that the DCC forecasted covariance matrices are superior to simpler models. We will therefore in the remaining part of this thesis apply the covariance matrix forecast produced by the DCC method.

	DCC - Rolling window		DCC- EWMA	
	Beta	t-statistic	Beta	t-statistic
Regional	-1.24E-05	-5.03**	-2.74E-05	-8.24**
Sector	-1.19E-05	-6.28**	-2.61E-05	-11.39**

Table 9: t-statistics for EM test. HAC T-values are reported, two stars indicate 95% significance level.

7.4 Stationarity

A time series $\{y_t\}$ is stationary, if its mean, variance and autocorrelations can be approximated as an average of a sufficiently long series of realizations. It is important to ensure that the series are stationary since the framework implemented here requires this. The time series are first converted by the logarithmic difference before the Augmented Dickey Fuller test where conducted on all series. All null hypotheses are rejected; hence we consider the diff-log return series stationary. For details concerning stationarity and the augmented Dickey Fuller test, see Appendix B.

8. Portfolio Optimization Results

We suspected that dividing our investment universe into sectors might give better possibilities for diversification than dividing the world into regions. From the correlation plots of returns (see table 2 and 4) this seems to not be the case. However it will be interesting to compare properties of portfolios calculated from the two investment universes. The previous chapter also suggested that we can get a forecast of expected return using ARMA models, that captures autocorrelation in returns i.e. the momentum effect, this seems to be stronger for sectors, than for regions. (see table 7 and 8).

8.1 The Minimum Variance and Max Sharpe Portfolios

We found that the max Sharpe portfolio outperformed the Regional rebalancing strategy for both the sector and regional investment universe when measured by both average return and risk adjusted return before transaction costs. However, after transaction costs the situation is turned around. The Max Sharpe portfolio suffers from very unstable portfolio weights, and thus when optimized daily this results in very high turnover and transaction costs which leads to negative returns. Our hypothesis was that by dividing the investment universe into sectors instead of region we could achieve lower portfolio risk, due to increasing cross country globalization. After observing that the different sectors were more correlated than the regions we thought that this hypothesis would be rejected. However the minimum variance portfolio for sectors exhibited the lowest variance. On the other hand the two investment universes are not directly comparable, because there are three more sectors than there are regions. We further notice that the minimum variance portfolios allocate a large part to indices which exhibit low unconditional variance, as for instance to North America and consumer staples. (see figure 12 and 13). The variance plot (figure 7 and 8) clearly shows that the minimum variance portfolio achieves lower risk throughout the out of sample period. The max Sharpe portfolio for sectors achieve only about one fourth of the return the regional max Sharpe portfolio offers, making a regional subdivision of the global investment universe the preferred method, at least measured by the risk return offered by the max Sharpe portfolio. CVaR is also considerably more negative for the max Sharpe portfolios,

indicating higher losses in the case where the 5% threshold is breached. Transaction costs are set to 0.258% as estimated in NBIM (2003).

	Regional Rebalancing	Sector		Regional	
		Min Variance	Max Sharpe	Min Variance	Max Sharpe
Pre cost annual mean	1.42 %	0.94 %	3.17 %	-0.32 %	12.85 %
Post cost annual mean	1.40 %	-11.73 %	-99.05 %	-9.03 %	-69.00 %
Annual Std.Dev	20.17 %	13.17 %	22.39 %	15.54 %	21.07 %
Annual Sharpe pre cost	0.03	0.00	0.10	-0.07	0.57
Annual Sharpe ex cost	0.03	-0.96	-4.43	-0.63	-3.28
Daily turnover ²	0.00	0.19	1.53	0.15	1.40
Skewness	0.00	-0.49	-0.52	-0.13	0.06
Kurtosis	10.40	11.53	9.25	9.44	11.17
CVaR	-3.38 %	-1.93 %	-3.38 %	-2.53 %	-3.17 %

Table 10: Summary statistics of daily out-of-sample returns for comparable portfolios over the period 04.01.2000-12.04.2012.

$$^2 \text{ Turnover} = \frac{1}{N} \sum_{t=1}^{T-1} \sum_{i=1}^N |w_{(i)t+1} - w_{(i)t}|$$

where $w_{(i)t+1}$ is the weight of asset i at the start of rebalancing period $t+1$, $w_{(i)t}$ is the weight of that asset before rebalancing at $t+1$, T is the total number of rebalancing periods (days).

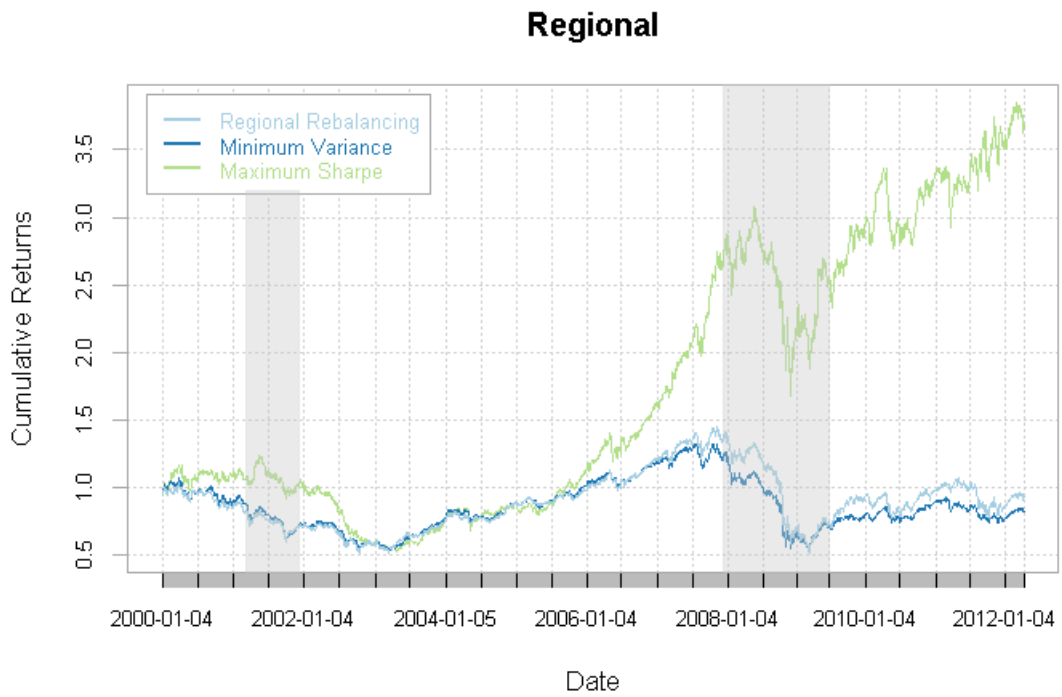


Figure 5: Cumulative portfolio returns before transaction costs over the period 04.01.2000-12.04.2012. Shaded area indicates a period of economic contraction.

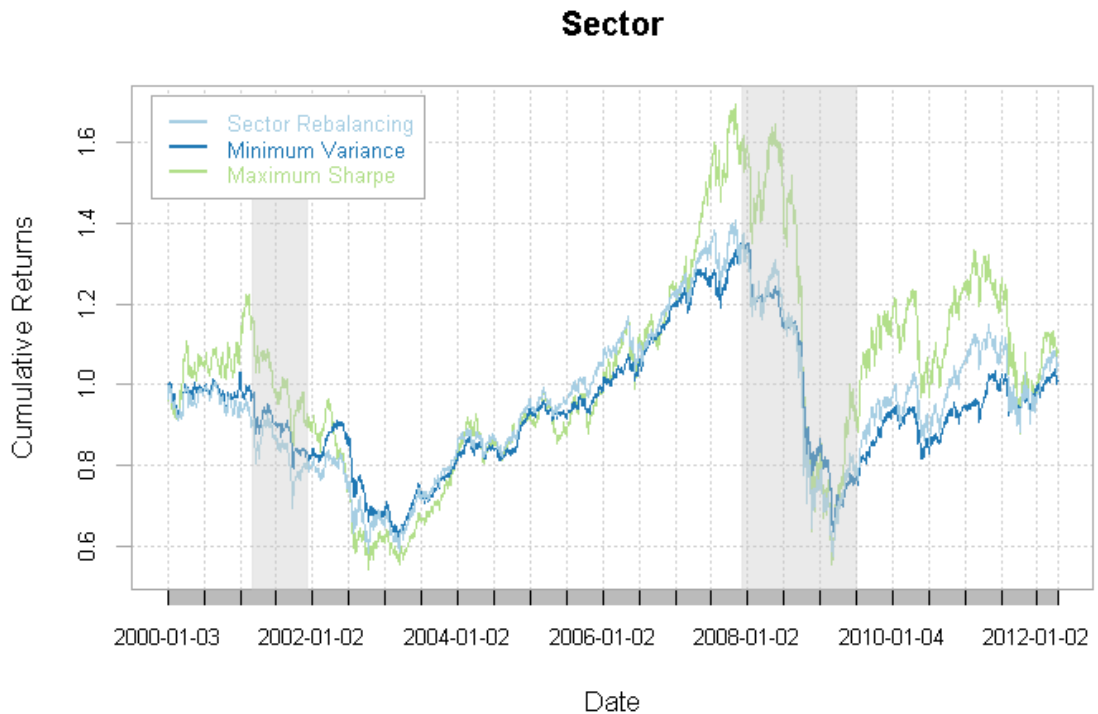


Figure 6 Cumulative portfolio returns before transaction costs over the period 04.01.2000-12.04.2012. Shaded area indicates a period of economic contraction.

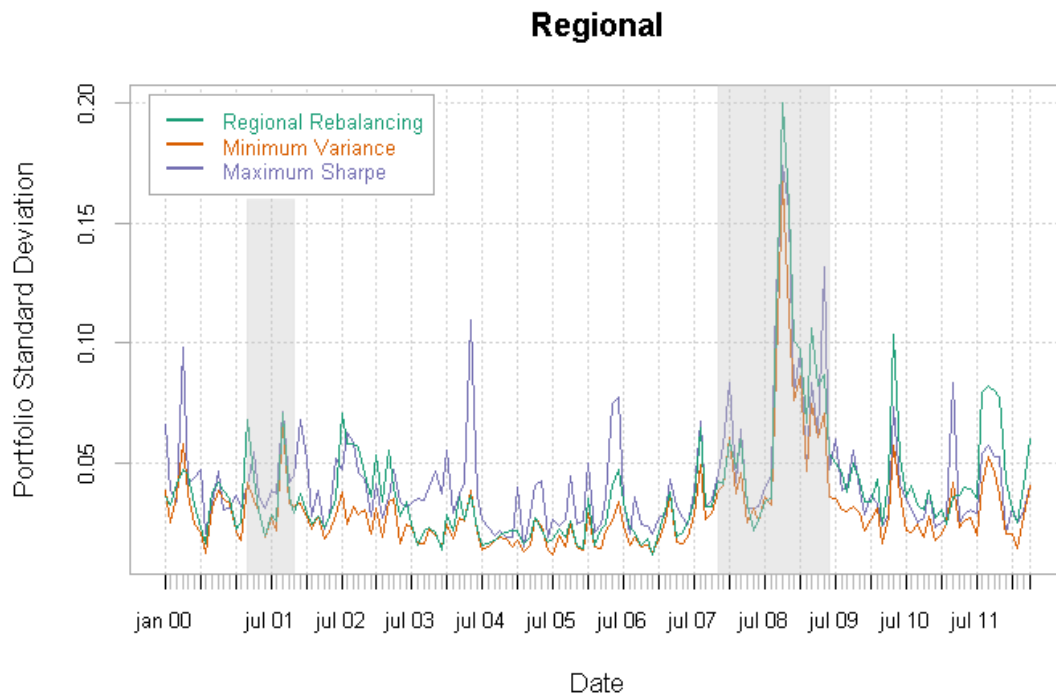


Figure 7: Monthly portfolio standard deviations over the period 04.01.2000-12.04.2012. Shaded area indicates a period of economic contraction.

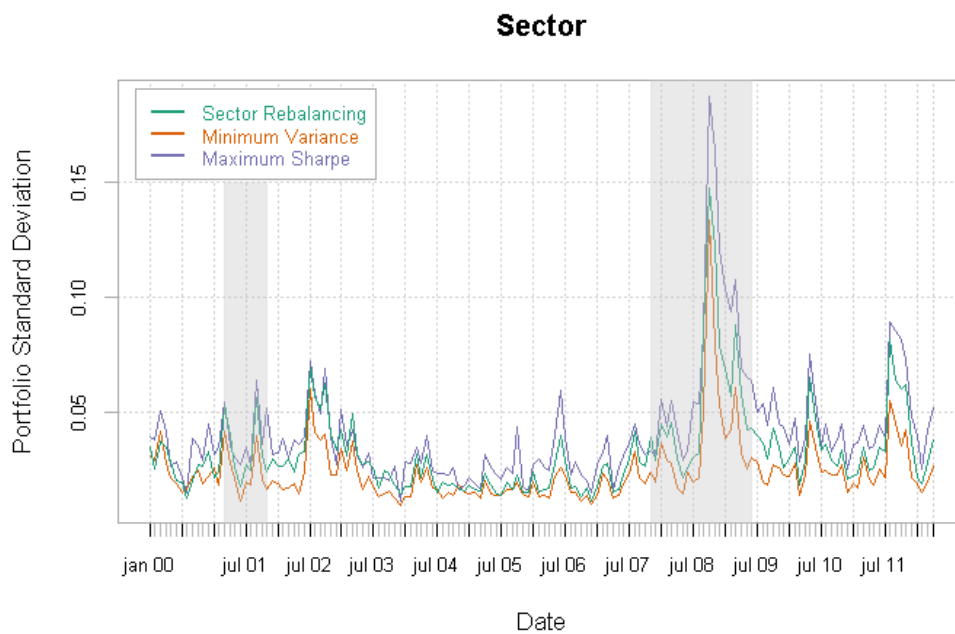


Figure 8: Monthly portfolio standard deviations over the period 04.01.2000-12.04.2012. Shaded area indicates a period of economic contraction.

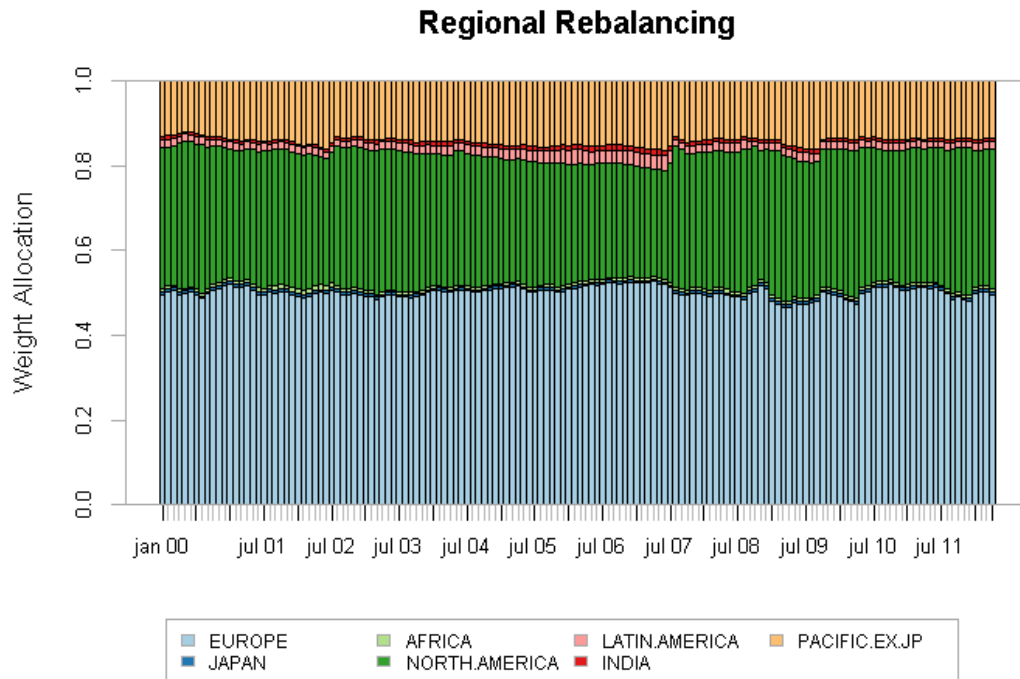


Figure 9: Stacked average monthly regional portfolio weights over the period 04.01.2000-12.04.2012.

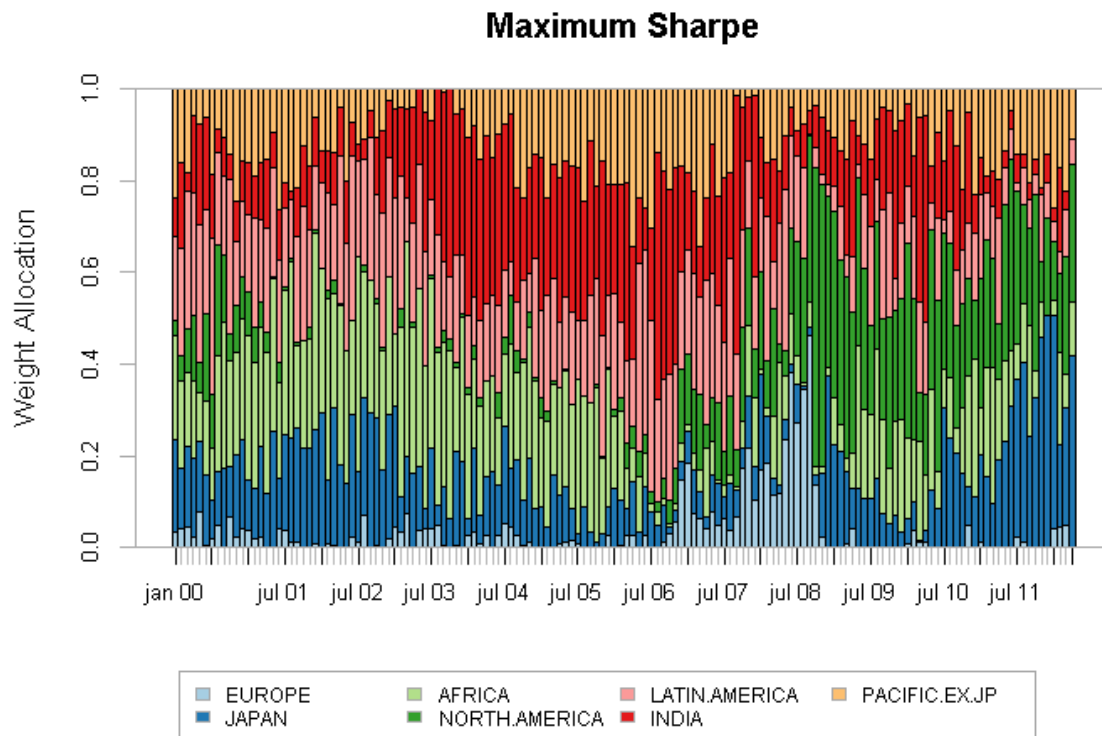


Figure 10: Stacked average monthly regional portfolio weights over the period 04.01.2000-12.04.2012.

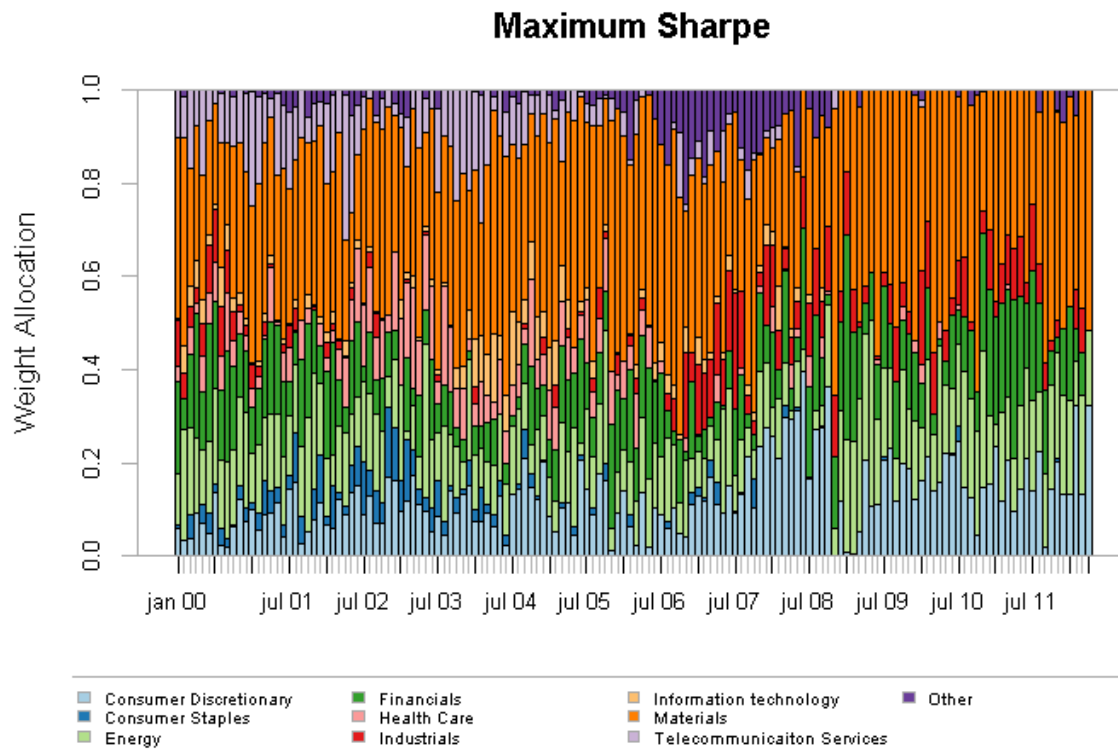


Figure 11: Stacked average monthly sector portfolio weights over the period 04.01.2000-12.04.2012.

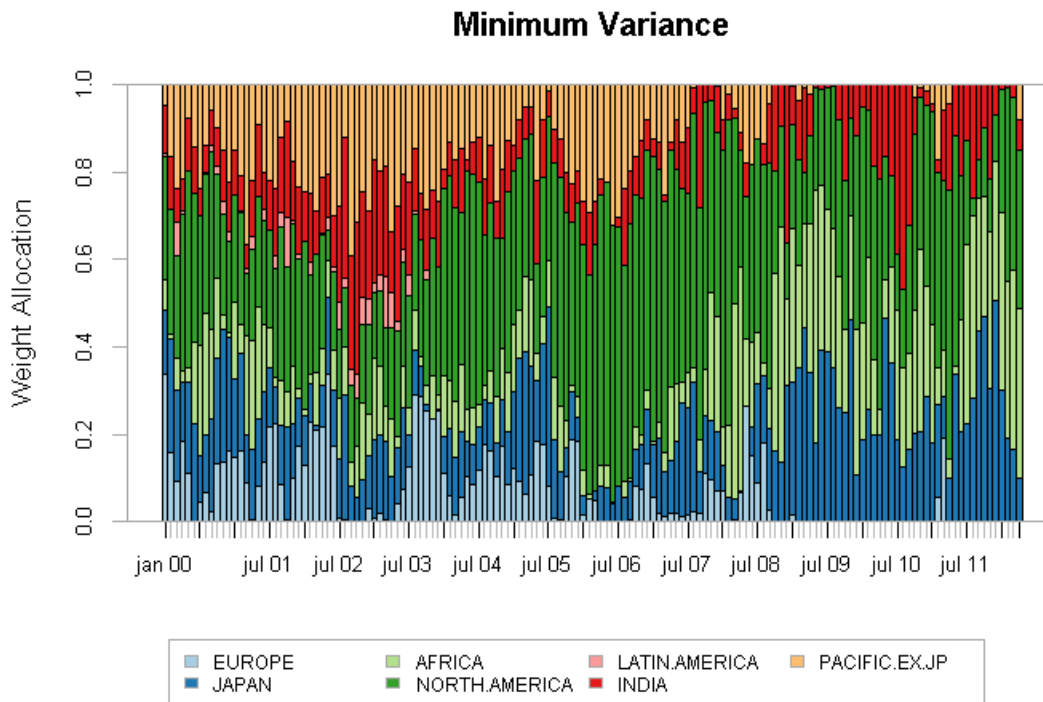


Figure 12: Stacked average monthly regional portfolio weights over the period 04.01.2000-12.04.2012.

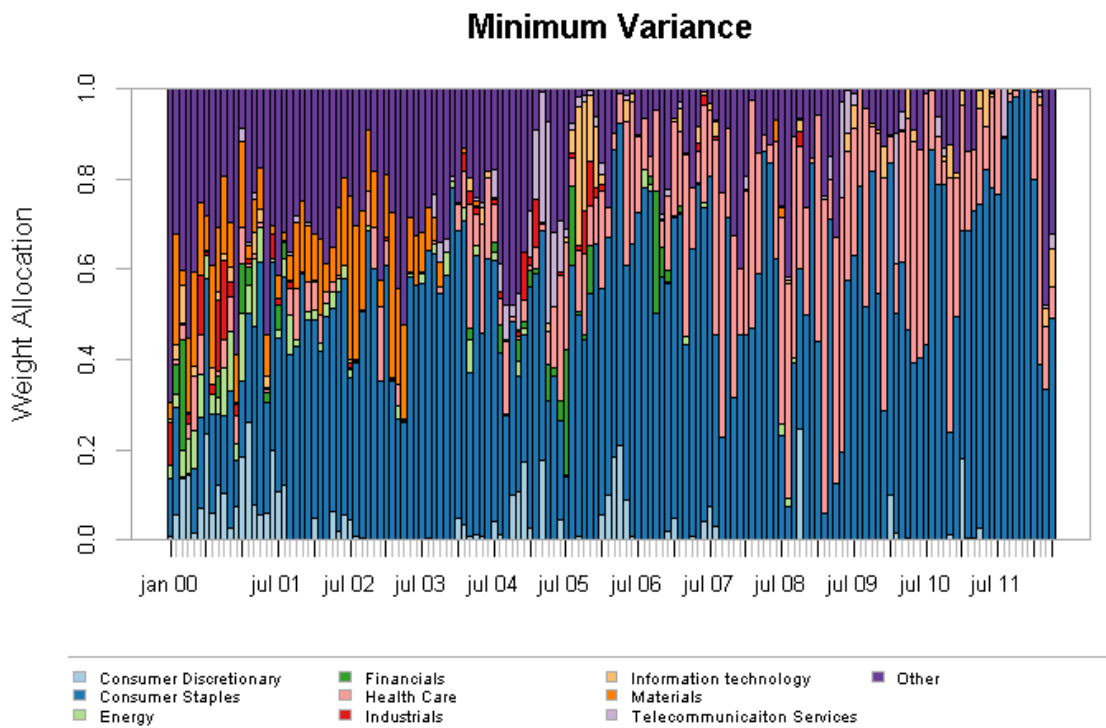


Figure 13: Stacked average monthly sector portfolio weights over the period 04.01.2000-12.04.2012.

8.2 Regional Portfolios with Additional Constraints

We found that changing the risk measure from standard deviation to CVaR had a marginal impact on the portfolio risk, measured by both standard deviation and CVaR for both the return target and risk budget portfolios. However the portfolios optimized with standard deviation as risk measure gave slightly higher returns and in turn higher risk adjusted return (Sharpe); this was true also after transaction costs. CVaR optimized portfolios had slightly lower kurtosis and skewness which were closer to zero. Optimizing with CVaR seems to give more unstable portfolio weights and thus higher turnover turnover. See table 11 and figure 15-18. Before trading costs both return target portfolios outperformed the regional rebalancing strategy but the regional rebalancing portfolio outperformed the risk budget portfolios. After trading costs however, the regional rebalancing portfolio again offered the best risk return payoff.

	Regional				
	Regional rebalancing	Return Target	Return Target CVaR	Std.Dev Risk Budgets	Risk Budgets CVaR
Pre cost annual mean	1.42 %	1.76 %	1.54 %	0.33 %	-0.16 %
Post cost annual mean	1.40 %	-26.58 %	-30.80 %	-1.32 %	-7.59 %
Annual Std.Dev	20.17 %	16.35 %	16.58 %	19.34 %	19.40 %
Annual Sharpe pre cost	0.032	0.06	0.05	-0.02	-0.05
Annual Sharpe ex cost	0.031	-1.67	-1.90	-0.11	-0.43
Daily turnover	0.00	0.48	0.55	0.03	0.13
Skewness	0.00	-0.56	-0.50	-0.04	-0.03
Kurtosis	10.40	10.95	10.15	11.15	10.98
CVaR	-3.38 %	-2.76 %	-2.71 %	-3.30 %	-3.32 %

Table 11: Summary statistics of daily out-of-sample returns for several portfolios over the period 04.01.2000-12.04.2012.

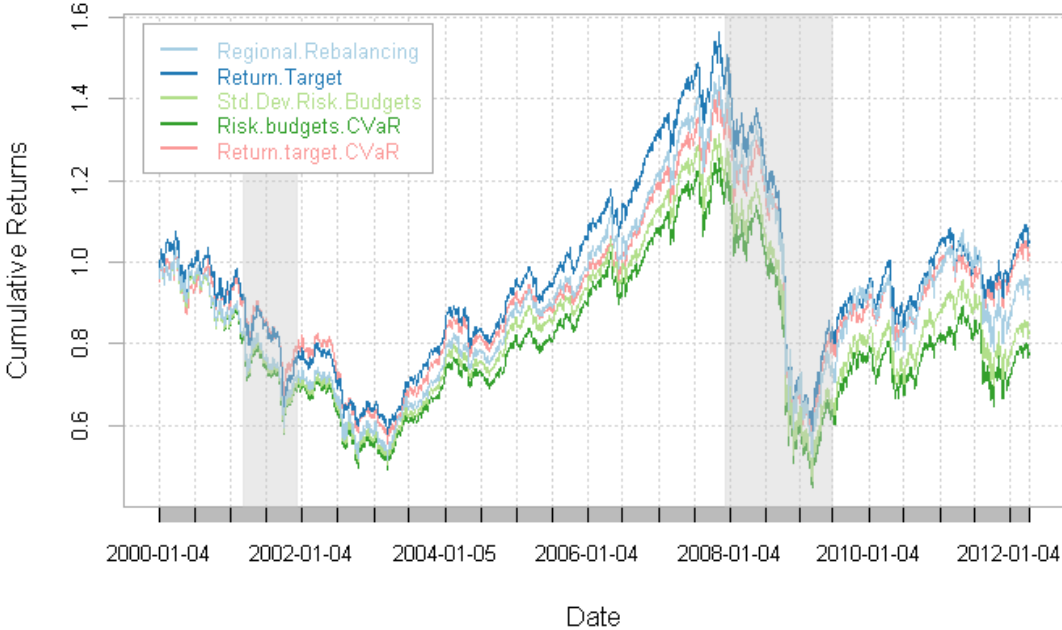


Figure 14: Cumulative portfolio returns before trading costs over the period 04.01.2000-12.04.2012. Shaded area indicates a period of economic contraction.

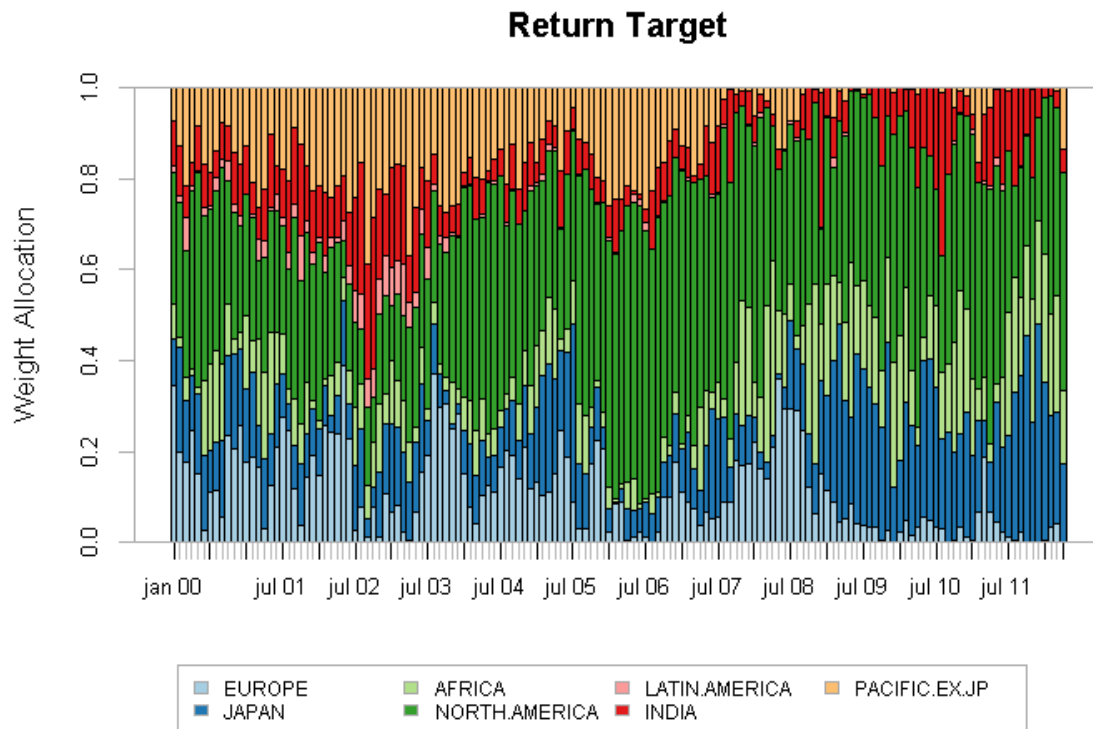


Figure 15: Stacked average monthly regional portfolio weights over the period 04.01.2000-12.04.2012.

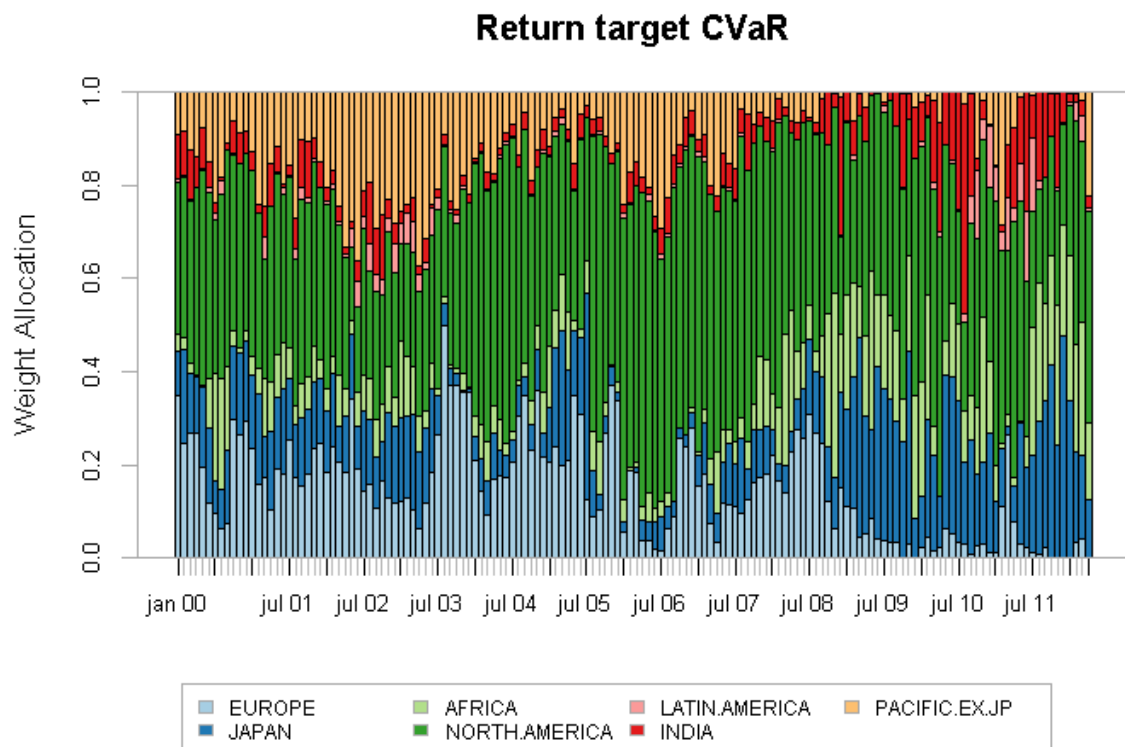


Figure 16: Stacked average monthly regional portfolio weights over the period 04.01.2000-12.04.2012.

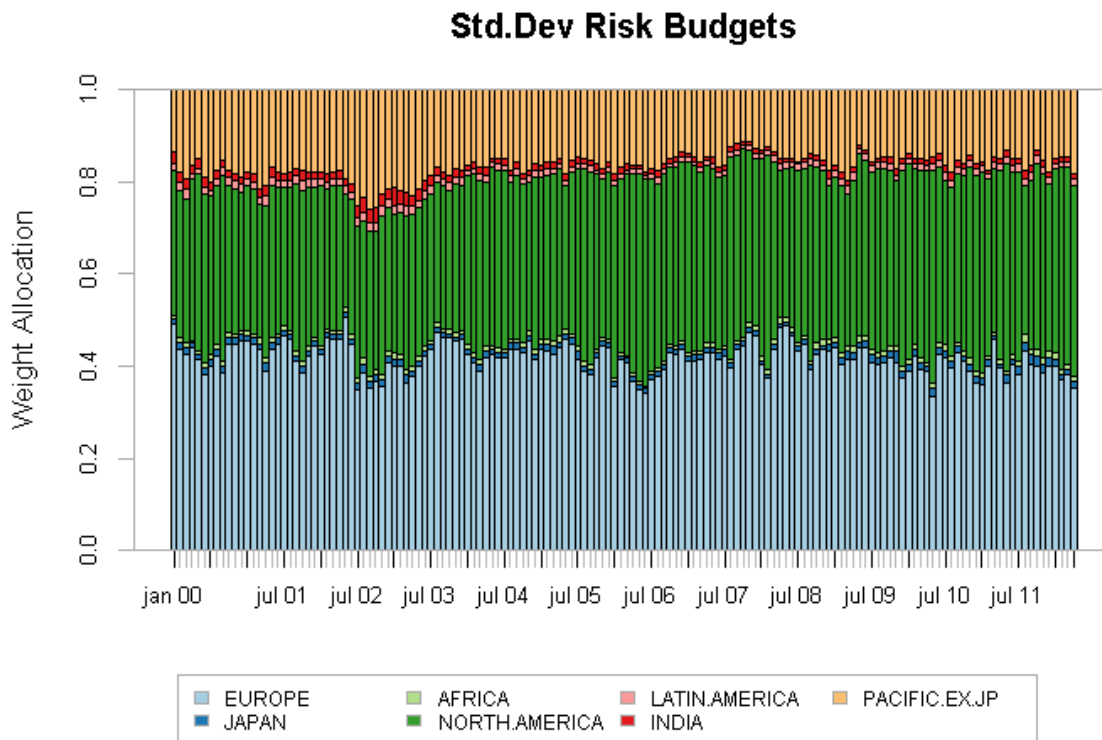


Figure 17: Stacked average monthly regional portfolio weights over the period 04.01.2000-12.04.2012.

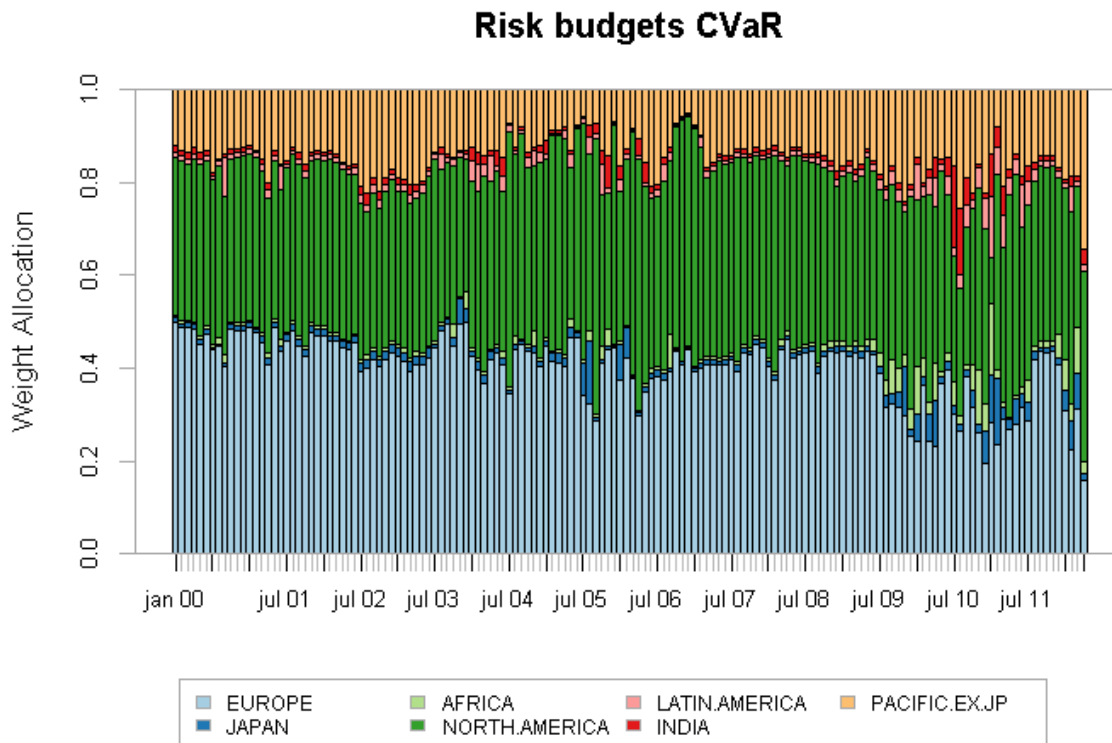


Figure 18: Stacked average monthly regional portfolio weights over the period 04.01.2000-12.04.2012.

8.3 Sector Results with Additional Constraints

We arrived at the same conclusion regarding the risk measure; changing it from standard deviation to CVaR has a marginal impact on portfolio characteristics. For instance one can look at figure 21 and 22 which clearly show close resemblance. In contrast to the regional results, the portfolios optimized with CVaR as risk measure instead of variance gave slightly higher returns and in turn higher risk adjusted return; this was true also after trading costs. The sector rebalancing portfolio outperformed all other portfolios measured by return and Sharpe ratio both before and after trading costs. We again notice a slight reduction in skewness and kurtosis and higher turnover with CVaR as risk measure.

	Sector				
	Sector Weights	Return Target	Return Target CVaR	Std.Dev Risk budgets	Risk Budgets CVaR
Pre cost annual mean	1.96 %	0.42 %	0.96 %	0.69 %	0.86 %
Post cost annual mean	1.94 %	-51.92 %	-55.14 %	-1.16 %	-6.61 %
Annual Std.Dev	17.62 %	14.30 %	14.50 %	16.57 %	16.84 %
Annual Sharpe pre cost	0.06	-0.03	0.00	-0.01	0.00
Annual Sharpe ex cost	0.06	-3.68	-3.85	-0.12	-0.45
Daily Turnover	0.00	0.78	0.84	0.03	0.11
Skewness	-0.30	-0.52	-0.50	-0.35	-0.33
Kurtosis	7.28	10.12	9.51	8.19	7.68
CVaR	-2.69 %	-2.18 %	-2.18 %	-2.56 %	-2.57 %

Table 12: Summary statistics of daily out-of-sample returns for several portfolios over the period 04.01.2000-12.04.2012.

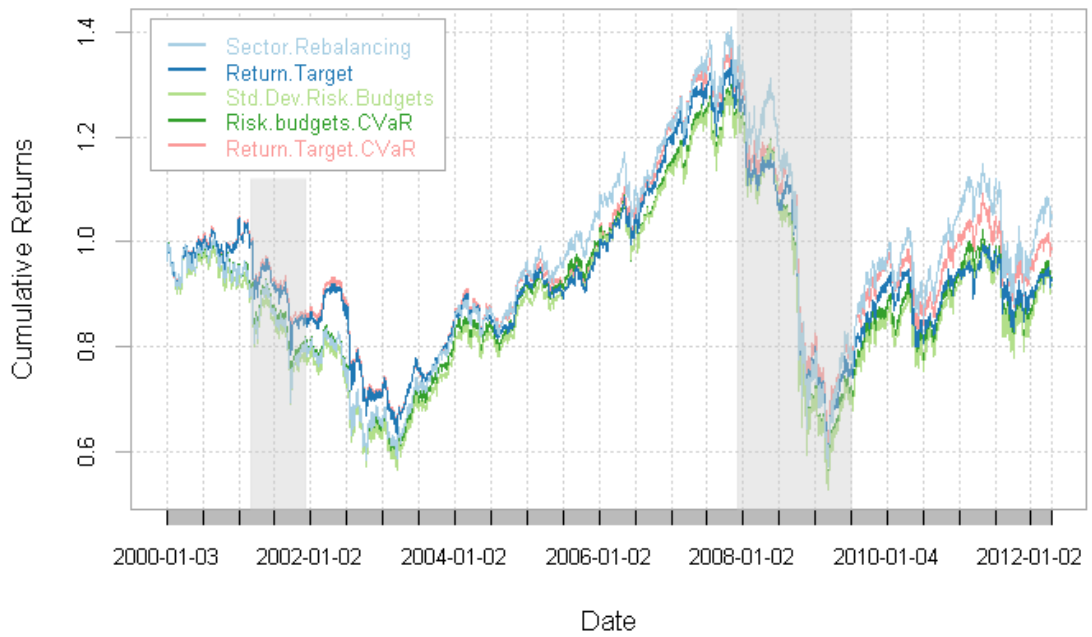


Figure 19: Cumulative portfolio returns over the period 04.01.2000-12.04.2012. Shaded area indicates a period of economic contraction.

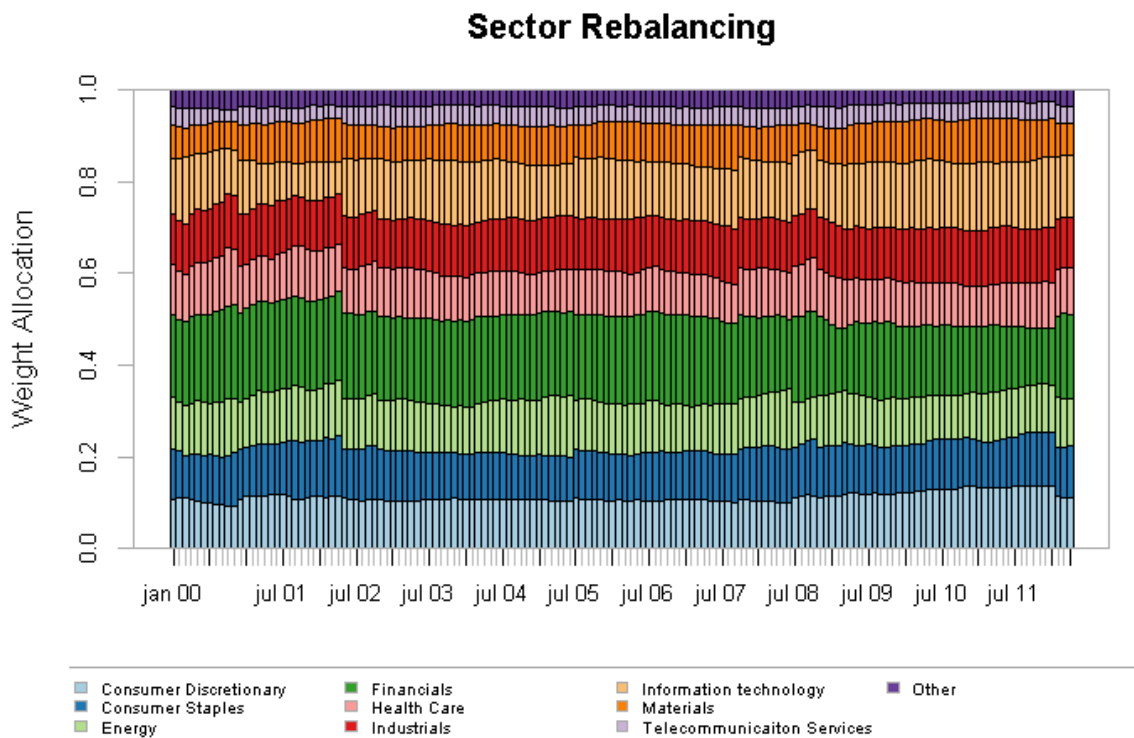


Figure 20: Stacked average monthly sector portfolio weights over the period 04.01.2000-12.04.2012.

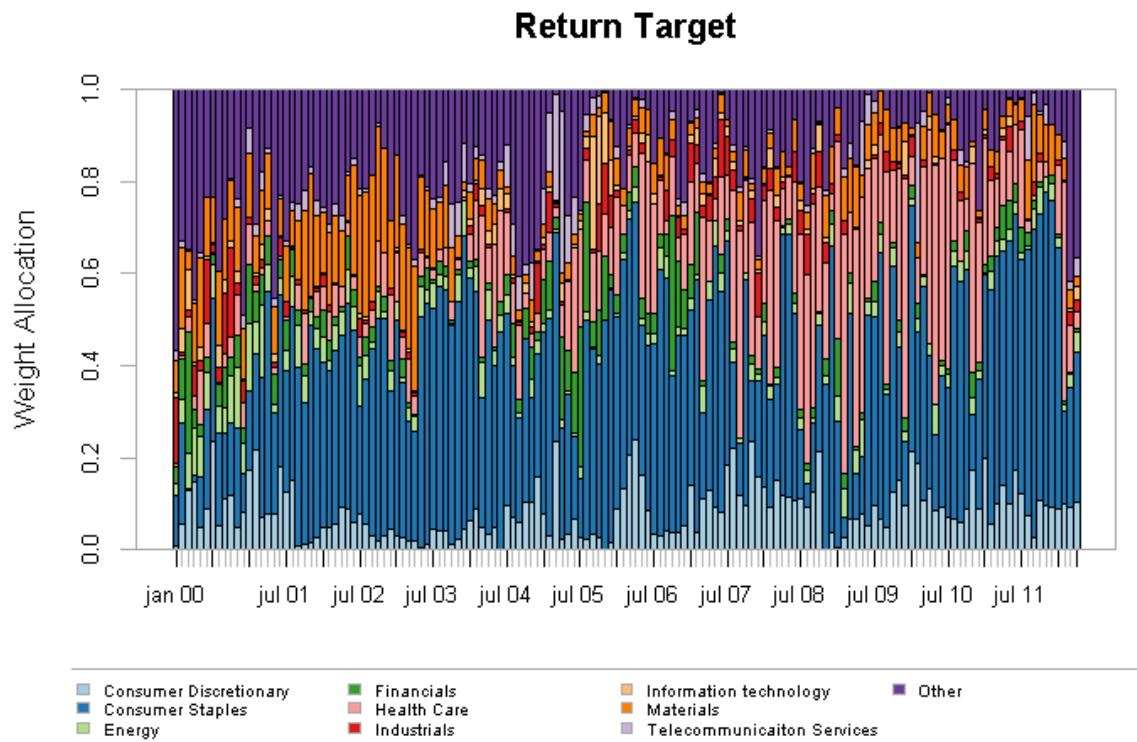


Figure 21: Stacked average monthly sector portfolio weights over the period 04.01.2000-12.04.2012.

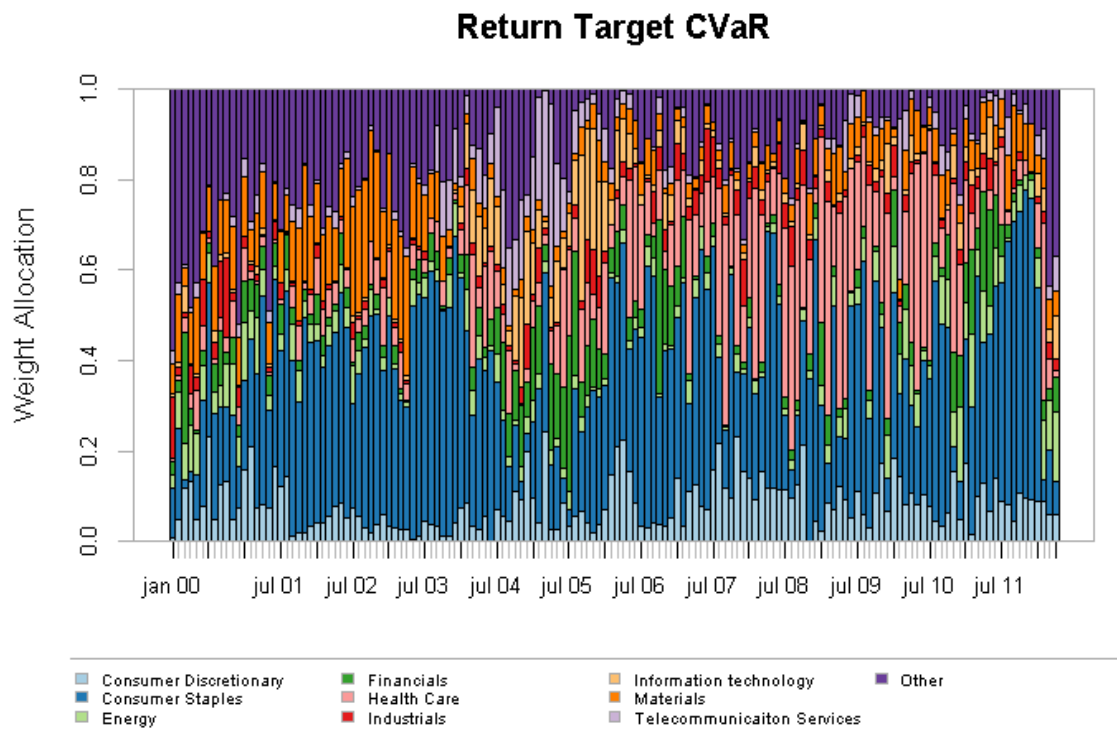


Figure 22: Stacked average monthly sector portfolio weights over the period 04.01.2000-12.04.2012.

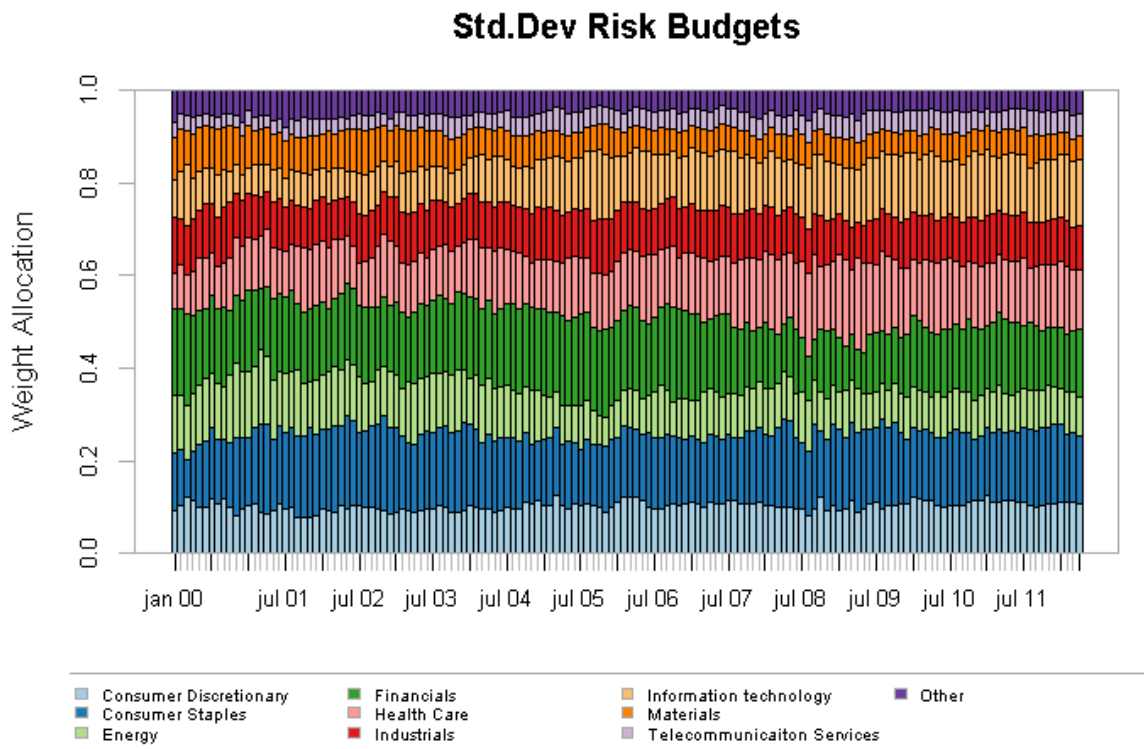


Figure 23: Stacked average monthly sector portfolio weights over the period 04.01.2000-12.04.2012.

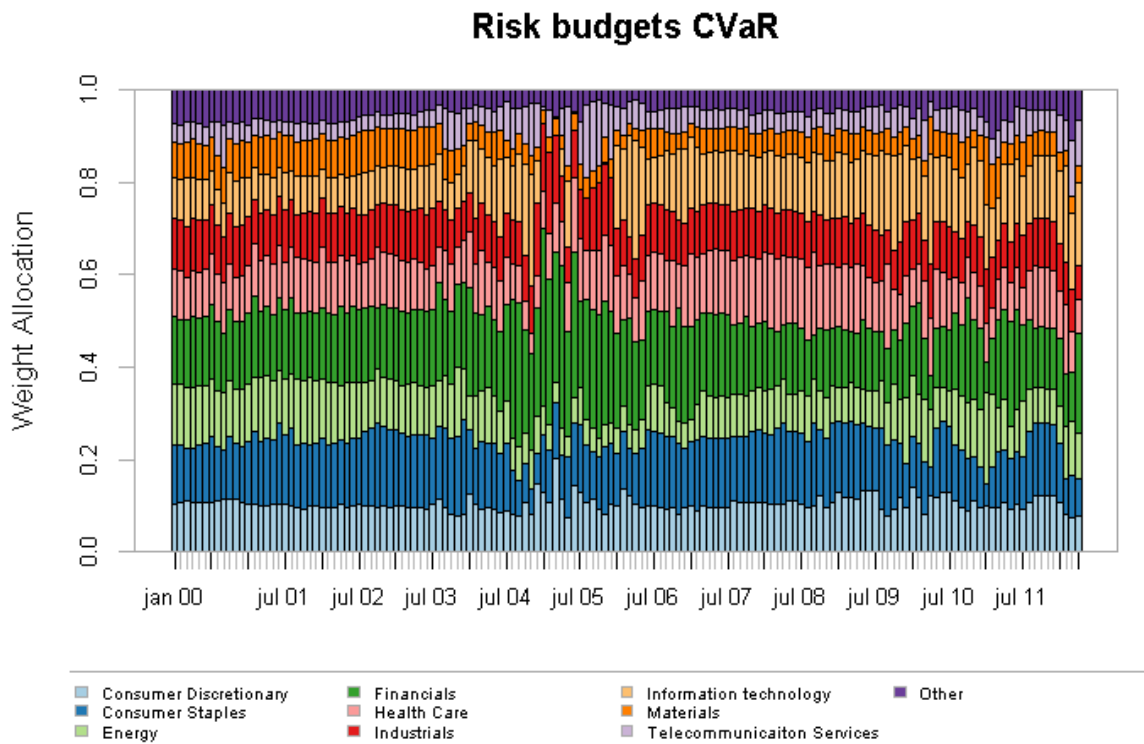


Figure 24: Stacked average monthly sector portfolio weights over the period 04.01.2000-12.04.2012.

9. The modified Regional Max Sharpe Portfolio

In part eight we found the max Sharpe regional portfolio to give the best pre cost results measured by the Sharpe ratio. However the portfolio suffered from an extremely high turnover which made the portfolio more of a theoretical exercise than a strategy suited for real life applications. Judging by the weight plot for maximum Sharpe (figure 10) a lot of persistence (autocorrelation) is present in the weights. It therefore seems sensible to smooth the series to create less volatile weights. To reduce the volatility in portfolio weights we calculated the portfolio weights as the mean of the optimal daily weights for 20-, 40-, 60- and 80 days in the past. We found this to substantially reduce the portfolio turnover and hence the trading costs, which in turn made this portfolio to outperform the regional rebalancing strategy also after transaction costs were imposed. Judging by figure 25 the reallocation during the financial crisis (fig 10 and 27) is less rapid, and some of the max Sharpe portfolio's benefit seems to have been lost. In table 13 it is interesting to notice that returns seems to increase with the number of days, and at the same time risk is increased, measured by CVaR and standard deviation. These portfolios also have higher kurtosis and are more skewed, indicating that they accumulate other kinds of risk than the regional rebalancing strategy. Figure 28 illustrates that the max Sharpe with 80 days smoothed weights average does not outperform the other portfolios during bull markets, but have consequently higher Sharpe ratio during bear markets. For SPU a max Sharp portfolio without further restrictions on regional weights would not be a possible equity allocation strategy. This is due to the size of the fund which would cause huge market impact costs in periods where the max Sharpe strategy allocates large parts of the capital to regions with relatively low market capitalization compared to the size of SPU. See figure 27.

	Regional Rebalancing	Max Sharpe	Max Sharpe 20	Max Sharpe 40	Max Sharpe 60	Max Sharpe 80
Pre cost mean	1.42 %	12.85 %	5.05 %	5.35 %	5.39 %	5.56 %
Post Cost Mean	1.40 %	-69.00 %	0.89 %	3.01 %	3.65 %	4.13 %
Standard Deviation	20.17 %	21.07 %	18.69 %	18.81 %	18.89 %	18.96 %
Sharpe pre Cost	0.0316	0.5729	0.2282	0.2428	0.2438	0.2522
Sharpe ex Cost (0.258%)	0.0308	-3.2804	0.0059	0.1182	0.1516	0.1762
Sharpe ex Cost (0.5%) ³	0.0301	-6.7678	-0.2022	0.0016	0.0653	0.1051
Daily turnover	0.000	1.3977	0.071	0.040	0.030	0.025
Skewness	-0.005	0.0623	-0.306	-0.290	-0.246	-0.239
Kurtosis	10.396	11.1680	10.465	11.313	11.654	11.633
CVaR	-3.38 %	-3.17 %	-2.97 %	-3.01 %	-3.07 %	-3.13 %

Table 13: Summary statistics of daily out-of-sample returns for realistic portfolios over the period 04.01.2000-12.04.2012.

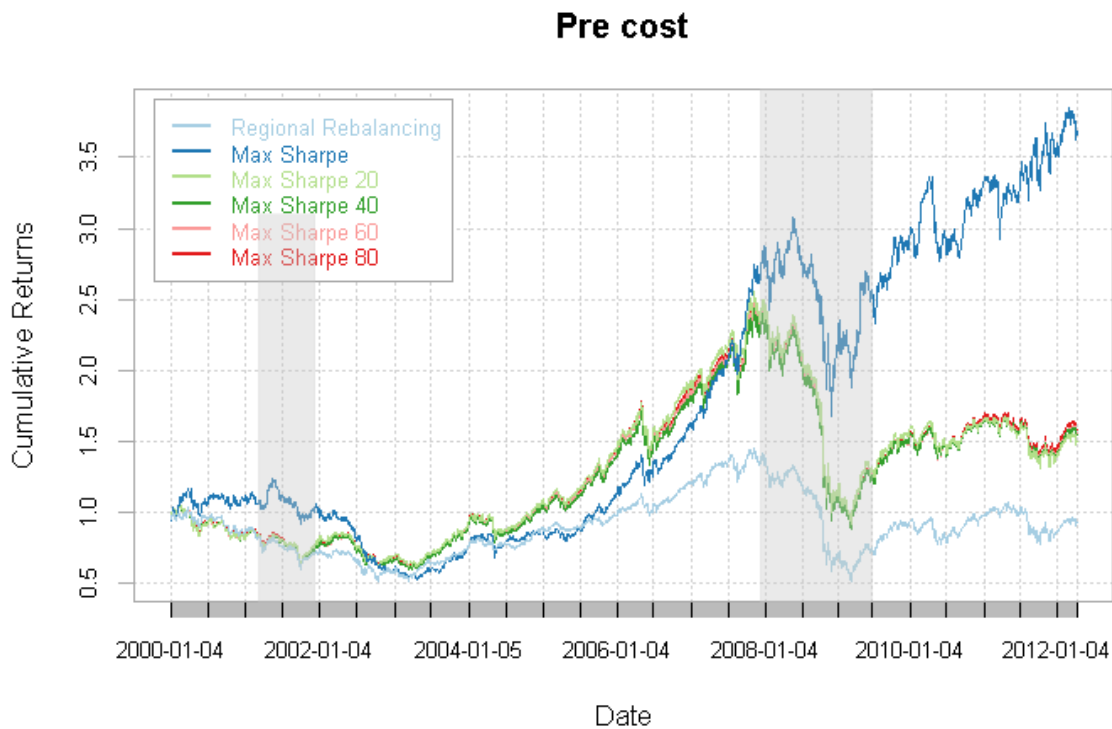


Figure 25: Cumulative portfolio returns over the period 04.01.2000-12.04.2012. Shaded area indicates a period of economic contraction.

³ Used in NBIM (2012) – Empirical Analysis of Rebalancing Strategies.

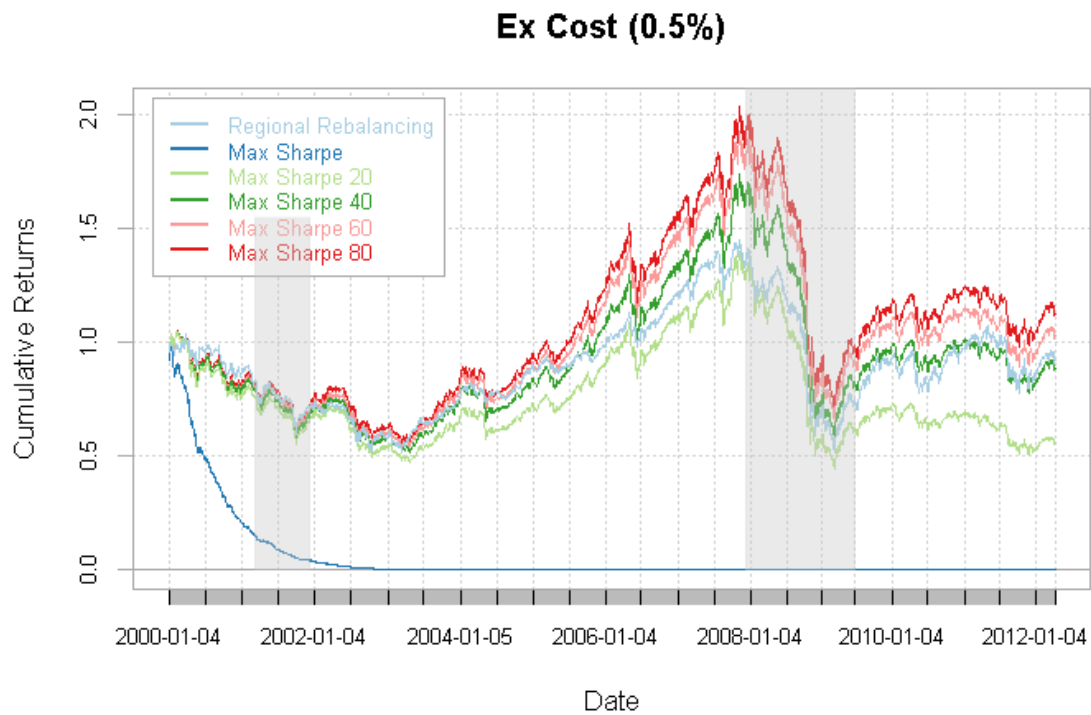


Figure 26: Monthly portfolio standard deviations over the period 04.01.2000-12.04.2012. Shaded area indicates a period of economic contraction.

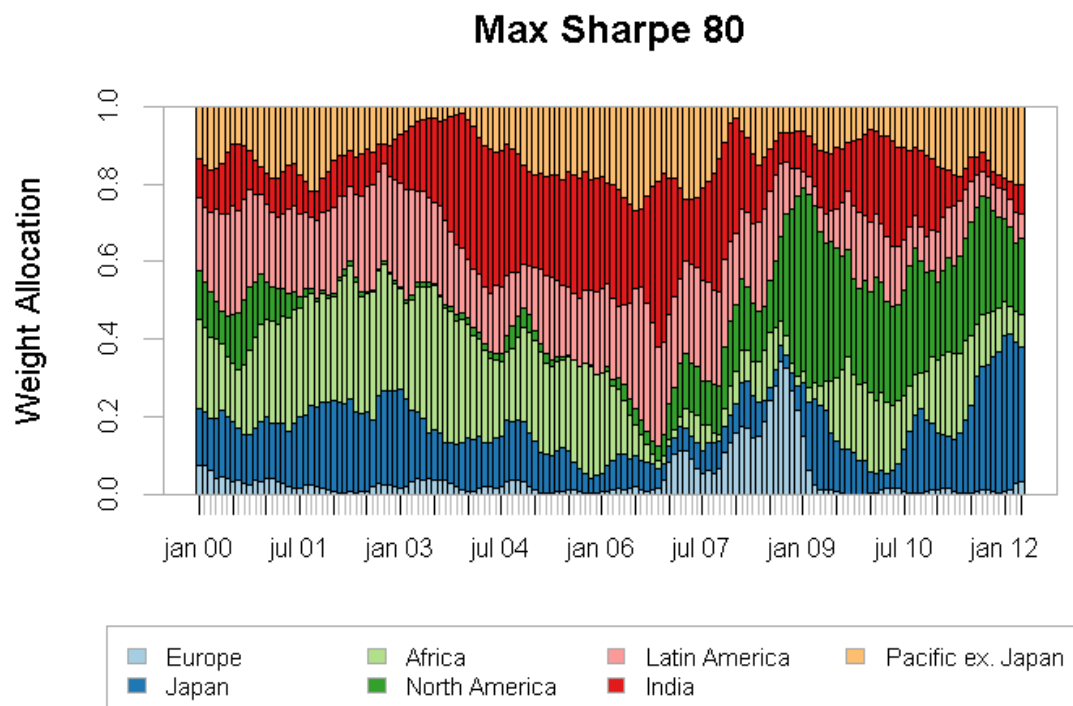


Figure 27: Stacked average monthly regional portfolio weights over the period 04.01.2000-12.04.2012.

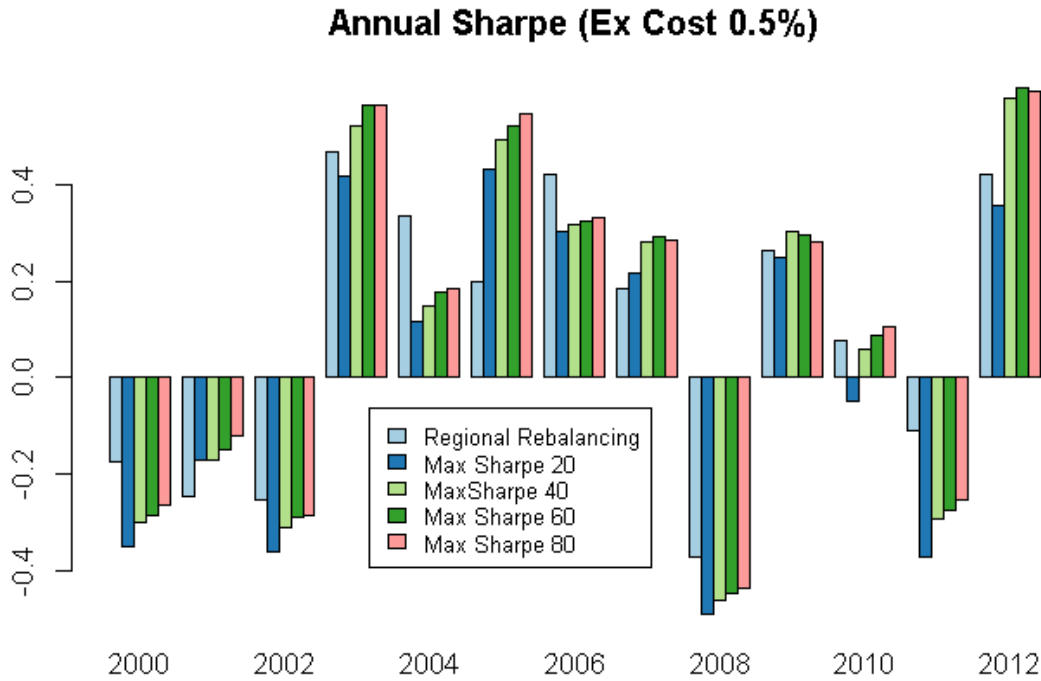


Figure 28: Annual Sharpe ratios ex cost (0,5%) over the period 2000-2012.

10. Conclusion

This thesis has evaluated the covariance forecasting ability of the moving average, EWMA and DCC model by comparing the unconstrained maximum Sharpe portfolios. We found the DCC model to give significantly better covariance forecasts than the other methods, and thus all portfolio optimization were based on covariance forecasts from the DCC method.

We further computed several portfolios based on both a sector and a regional subdivision of the investment universe and found that the main differences were that a sector subdivision produced the minimum variance portfolio with the lowest variance, but the regional subdivision gave a much better max Sharpe portfolio. For both investment universes, changing the risk measure from variance to CVaR had a marginal impact on the out of sample performance.

The regional max Sharpe portfolio exhibited the highest mean and risk return payoff, but suffered from extremely high turnover. After modifying the portfolio by determining the portfolio weights as an average of optimal portfolio weights in a given number of days in the past, this asset allocation outperformed the regional rebalancing portfolio also after transaction costs.

Based on our results we recommend optimizing the asset allocation according to a maximum Sharpe objective with a regional subdivision of the global investment universe combined with a method of reducing the transaction costs. However, to reduce market impact costs this requires that the amount of capital under management is relatively small compared to the market capitalization of the regions invested in.

Further research with longer forecasting horizon and out of sample period would be an interesting extension of our thesis. It would also be interesting to evaluate different methods of creating stable weights for the regional max Sharpe portfolio.

11. References

(2012) <http://www.treasury.gov/resource-center/data-chart-center/interest-rates/Pages/TextView.aspx?data=yield>

Ang, A. and G. Bekaert. (1999) *International Asset Allocation with Time-Varying Correlations*. Working Thesiss 7056, National Bureau of Economic Research.

Ang, A. Goetzmann, W. Schaefer, S (2009). *Evaluation of Active Management of the Norwegian Government Pension Fund – Global*,

Ardia et al (2011) *Large-scale portfolio optimization with DEoptim*.

Artzner, P. Delbaen, F. Eber, J. and Heath, D (1999) *Coherent Measures of Risk*. *Mathematical Finance*, 9, 203-228.

Benartzi, S and Thaler, R (1995) *Myopic loss aversion and the equity premium puzzle*. *The Quarterly Journal of Economics*, Vol. 110, No. 1 pp. 73-92.

Boudt, K and Peterson, B (2009)
https://rforge.rproject.org/scm/viewvc.php/trunk/trunk/R/Risk_budget_functions.R?view=markup&root=returnanalytics&pathrev=1378

Boudt, K et al (2011) *Asset Allocation with Conditional Value-at-Risk Budgets*.

Boudt, K. Peterson, B. and Croux, C (2008) *Estimation and decomposition of downside risk for portfolios with non- normal returns*. *Journal of Risk*, Winter 2008, 79 –103.

Cao, Z. Harris, R. Shen, J (2009) *Hedging and Value at Risk: A Semi-Parametric Approach*. *The Journal of Futures Markets*. Vol 30, Number. 8.

Carl, P and Peterson, B (2010) *Performance Analytics an econometric tool for performance and risk analysis*. R package

Clive, W. Granger, M and Machina, J (2005) *Structural attribution of observed volatility clustering*.

Cochrane, J (1999) *New Facts in Finance*, NBER Working Paper No. 7169

Cornish, E. Fisher, R (1938) *Moments and Cumulants in the specification of distributions*.
Revue de l'institut International de statistique 5. 1937. 307-322

Daniélsson, J (2011) *Financial Risk Forecasting*.

Danthine, J and Donaldson, J (2005) *Intermediate Financial Theory*, 2nd ed., Elsevier Academic Press.

Das, S. and Uppal, R (2001) *Systemic risk and international portfolio choice*. Manuscript, Harvard University.

DeMiquel, V. Garlappi, L and Uppal, R (2009) *Optimal versus naïve diversification: how inefficient is the 1/N portfolio strategy?* Review of financial studies 22, 1915-1953.

Diebold, F and Roberto, S (1995) *Comparing Predictive Accuracy*. Journal of Business & Economic Statistics 13, 253-263.

Engle, R (1982) *Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation*. Econometrica, 50, 987–1006.

Engle, R (2002) *Dynamic conditional correlation – a simple class of multivariate GARCH models*.

Engle, R and Colacito, R (2006) *Testing and Valuing Dynamic Correlations for Asset Allocation*, Journal of Business & Economic Statistics 24, 238-253

Engle, R and Sheppard, K (2001) *Theoretical and Empirical properties of Dynamic Conditional Correlation Multivariate GARCH*.

Fama, E and French, K (1996) *Multifactor Explanations of Asset Pricing Anomalies*. The Journal of Finance. Vol. 51. No. 1 p. 55-84

Fama, E and MacBeth, J (1973) *"Risk, Return, and Equilibrium: Empirical Tests."* Journal of Political Economy.

Figlewski, S and Wang, X (2000) *Is the leverage effect a leverage effect?* Working thesis

Gibbons, M. Ross, S and Shanken, J (1989) “*A Test of the Efficiency of a Given Portfolio*”, *Econo-metrica*, 57, 5, 1121-1152.

Hanssens, D. Parsons, J and Schultz, L (1990) *Market response Models*. Econometric and time series analysis. Page 143

Hens, T and Steude, S (2009) *The leverage effect without leverage*.

J.P. Morgan (1994) *RiskMetrics*. Second Edition.

J.P. Morgan. (1996) *RiskMetrics – Technical Document*. Fourth Edition. 78-80

Jagannathan, R and Ma, T (2003) *Risk reduction in large portfolios: why imposing the wrong constraints helps*. Working paper 8922 <http://www.nber.org/papers/w8922>

Kolusheva, D (2008) *Out-of-sample Performance of Asset Allocation Strategies*.

Longin, F and B, Solnik (2002) *Extreme Correlation of International Equity Markets*. The Journal of Finance.

Lundbergh, S and Teräsvirta, T (2002) *Evaluating GARCH models*. Journal of Econometrics 110. 417-435

Maillard, S. Roncalli, T. and J. Teiletche (2010) *On the properties of equally-weighted risk contributions portfolios*. Journal of Portfolio Management , Summer 2010 , P. 60 – 70

Markowitz, H (1952) *Portfolio Selection*. Journal of finance, 7(1), 77–91.

Mausser, H and Rosen, D (1999) *Beyond VaR: From measuring risk to managing risk*. *ALGO Research Quarterly*. Vol, 2, 5-20

Merton, R (1980) *On estimating the expected return on the market: An exploratory investigation*, Journal of Financial Economics, Vol 8, 323-361.

NBER (2012) <http://www.nber.org/cycles.html>

NBIM (2003) *Costs associated with large equity trades*.

NBIM (2011) <http://www.nbim.no/Investeringer/sammensetning/referanseindeksene/1058/>

NBIM (2012) *Alternatives to Market-value-weighted index*. (Discussion Note)

NBIM (2012) *Empirical analysis of rebalancing strategies*. NBIM discussion note number 3 2012.

Norges Bank. (2012) <http://www.nbim.no/no/media-og-publikasjoner/Brev-til-Finansdepartementet/2012/statens-pensjonsfond-utland-regler-for-rebalansering-av-referanseindeksen/>. Letter to NBIM.

Patton, A and Sheppard, K (2007) *Evaluating volatility and correlation forecasts*.

Price, K. Storn, R and Lampinen, J (2006) *Differential Evolution - A Practical Approach to Global Optimization*. Berlin Heidelberg: Springer-Verlag. ISBN 3540209506

Qian, E (2006) *On the financial interpretation of risk contributions: Risk Budgets do add up*, Journal of Investment Management, Fourt Quarter.

Reuters (14. August 2009) <http://www.reuters.com/article/2009/08/14/norway-oilfund-idUSOSN00233820090814>

Rockafellar, T and Uryasev, S (2000) *Optimization of Conditional Value-at-risk*. *Journal of Risk*. Vol 2, Number 3.

Scowcroft, A and Sefton, J (2005) *Understanding momentum* Financial Analyst Journal, Vol 61, No 2, pp. 64-82

Sharpe, W (1964) *A Theory of Market Equilibrium under Conditions of Risk*. The Journal of Finance. Vol 19. No 3. pp. 425-442.

Storn, R and Price, K (1997) *Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces*. Journal of Global Optimization 11: 341–359.

12. Appendix

12.1 Appendix A: Data Cleaning

Portfolio moments and especially higher moments are extremely sensitive to data spikes, for this reason we will “clean” the data in the CVaR portfolio optimization, using the cleaning method “Boudt”. This is a robust method that does not remove data from the series, but only decreases the magnitude of extreme events. When estimating the downside risk with loss probability α , observations that belongs to the $(1 - \alpha)$ last observations will not be cleaned.

The time series is cleaned in three steps as specified in Peterson et al. (2010). First suppose that we have an n -dimensional vector time series of length T : r_1, \dots, r_T .

1. Ranking the observations based on their extremeness: Let μ and Σ is the mean and covariance matrix of the bulk of the data and let $[\cdot]$ be the operator that takes the integer part of its argument. The squared Mahalanobis distance $d_t^2 = (r_t - \mu)' \Sigma^{-1} (r_t - \mu)$ is used as a measure of the extremeness of the return observation r_t . μ and Σ are estimated as the mean and covariance matrix of the subset size $[(1 - \alpha)T]$ for which the determinant of the covariance matrix of the elements in that subset is the smallest. These estimates are then robust against the α most extreme returns. Let $d_{(i)}^2, \dots, d_{(T)}^2$ be the ordered sequence of the estimated squared Mahalanobis distances such that $d_{(i)}^2 \leq d_{(i+1)}^2$.
2. Outlier identification: Returns are categorized as outliers if their estimated squared Mahalanobis distance d_t^2 is greater than the $(1 - \alpha)$ quantile $d_{((1-\alpha)T)}^2$ and exceeds an extreme quantile of the Chi-square distribution function with n degrees of freedom, which is the distribution function of d_t^2 when the returns are normally distributed. In this application the 99,9% quantile is denoted $\chi_{n,0.999}^2$.
3. Data cleaning: The returns r_t that are identified as outliers in step 2 are replaced by:

$$r_t \sqrt{\left(\frac{\max(d_{((1-\alpha)T)}^2, \chi_{n,0.999}^2)}{d_t^2} \right)}$$

The cleaned return vector has the same orientation as the original return vector, but its magnitude is smaller.

12.2 Appendix B: Stationarity

If a time series $\{y_t\}$ is a stationary series, this means that the mean, variance and autocorrelations can be approximated as an average of sufficiently long series of realizations. A stochastic process having a finite mean and variance is covariance stationary for all t and $t - s$:

$$\begin{aligned} E(y_t) &= E(y_{t-s}) = \mu \\ E[(y_t - \mu)^2] &= E[(y_{t-s} - \mu)^2] = \sigma_y^2 \quad [var(y_t) = var(y_{t-s}) = \sigma_y^2] \\ E[(y_t - \mu)(y_{t-s} - \mu)] &= E[(y_{t-j} - \mu)(y_{t-j-s} - \mu)] = \gamma_s \\ [cov(y_t, y_{t-s}) &= cov(y_{t-j}, y_{t-j-s}) = \gamma_s] \end{aligned}$$

γ_0 is equivalent to the variance of y_t . A time series is thus covariance stationary if its mean and all autocovariances are unaffected by a change in time origin. In this thesis we will only consider weakly stationary series, so a series denoted as stationary means that the series is a weakly stationary series.

Test of stationarity: If the series contain a unit root an ARMA model will not fit the dataset, and thus we will have to make the series stationary. To test this we applied the Dickey-Fuller test (Dickey and Fuller (1979)) this is a test used to find out whether a time series contains a unit root. We applied the extended version called the augmented Dickey-Fuller (ADF) test, which for many time series gives a better fit than the simpler version of the test. The appropriate number of differentiated lags was decided based on Aikaikes information criteria.

$$\begin{aligned} \Delta y_t &= \gamma y_{t-1} + \sum_{i=2}^p \beta_i \Delta y_{t-i+1} + \varepsilon_t \\ \Delta y_t &= \alpha_0 + \gamma y_{t-1} + \sum_{i=2}^p \beta_i \Delta y_{t-i+1} + \varepsilon_t \\ \Delta y_t &= \alpha_0 + \gamma y_{t-1} + \alpha_2 t + \sum_{i=2}^p \beta_i \Delta y_{t-i+1} + \varepsilon_t \end{aligned}$$

The difference between the three regressions is the deterministic elements α_0 and $\alpha_2 t$. The null hypothesis is $\gamma = 0$, γ is defined as $(\alpha_1 - 1)$, hence the null hypothesis is that the time series contains an unit root. Under null hypothesis $\gamma = 0$, the first regression is a random walk model, the second with a drift term, and the third includes both a drift and a linear time trend. Each specification of the test has its own test statistics, denoted $\tau, \tau_\mu, \tau_\tau$ respectively. The time series are first converted by the logarithmic difference before the test specified without a drift or a trend term where conducted, all null hypothesis could be rejected, hence we consider the diff-log return series stationary.

	Europe	Japan	Africa	North America	Latin America	India	Pacific ex. Japan
T-statistics	-46	-47	-44	-46	-44	-43	-45

Table 14: T-values from ADF test, conducted at regions at the entire sample period.

	Consumer Discretionary	Consumer Staples	Energy	Financials	Health Care
T-statistics	-46	-49	-50	-45	-48

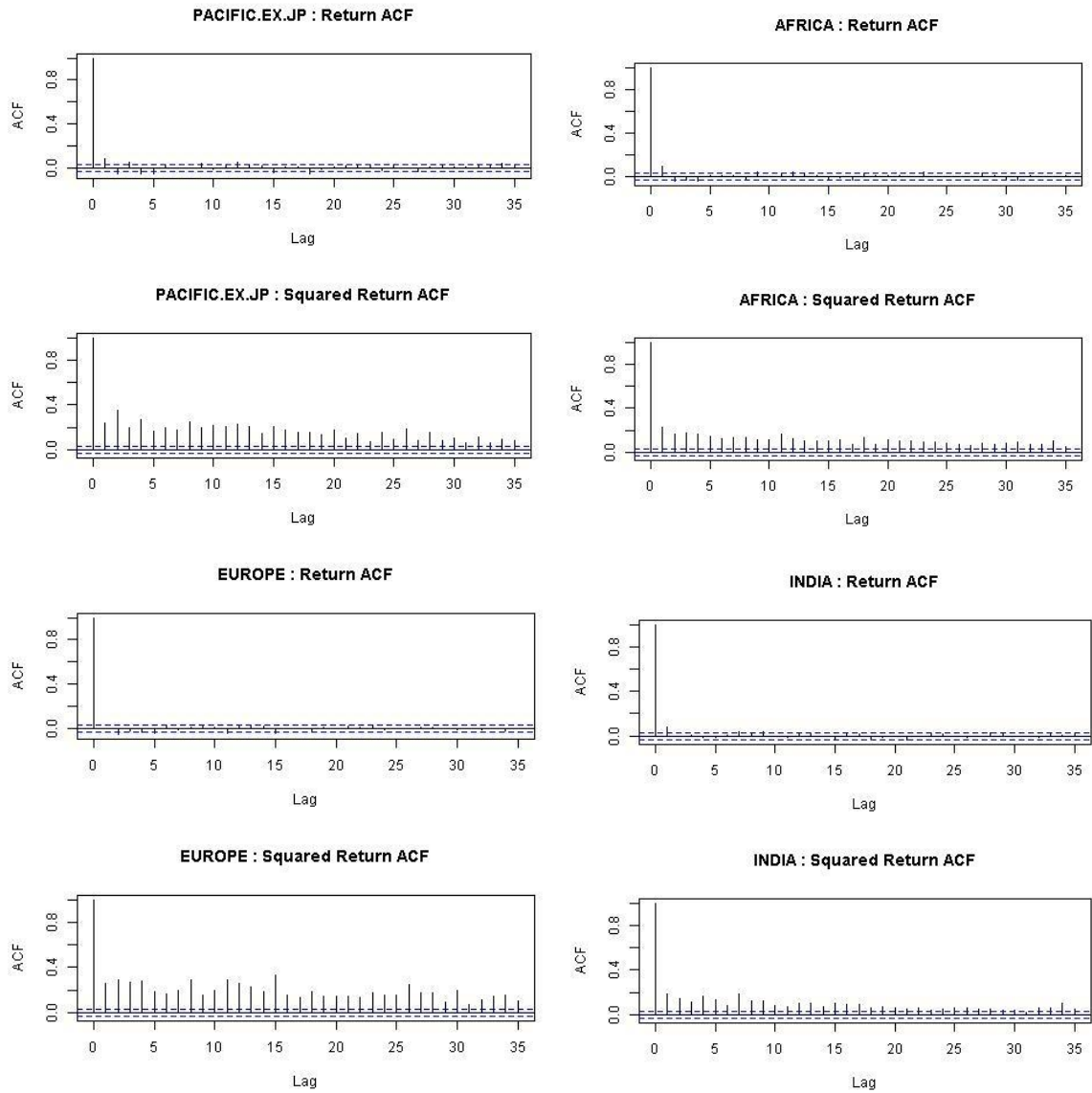
	Industrials	Information technology	Materials	Telecommunication Services	Other
T-statistics	-46	-47	-46	-48	-49

Table 15: T-values from ADF test, conducted at sectors for the entire sample period.

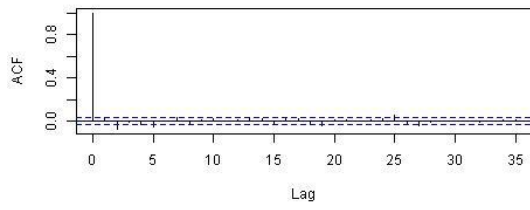
12.3 Appendix C: Diagnostic Plots

12.3.1 Regions

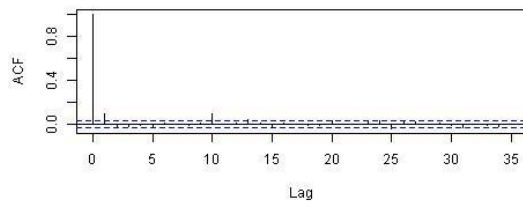
Original series ACF and ACF for squared return:



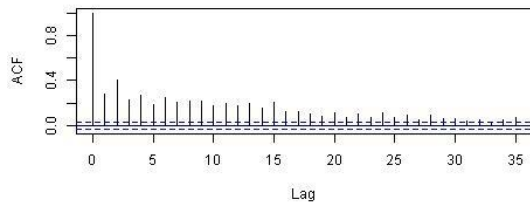
JAPAN : Return ACF



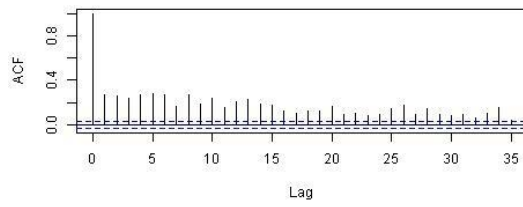
LATIN.AMERICA : Return ACF



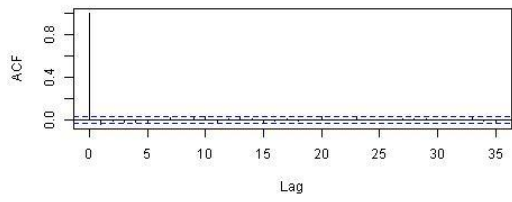
JAPAN : Squared Return ACF



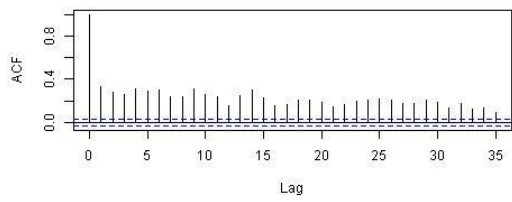
LATIN.AMERICA : Squared Return ACF



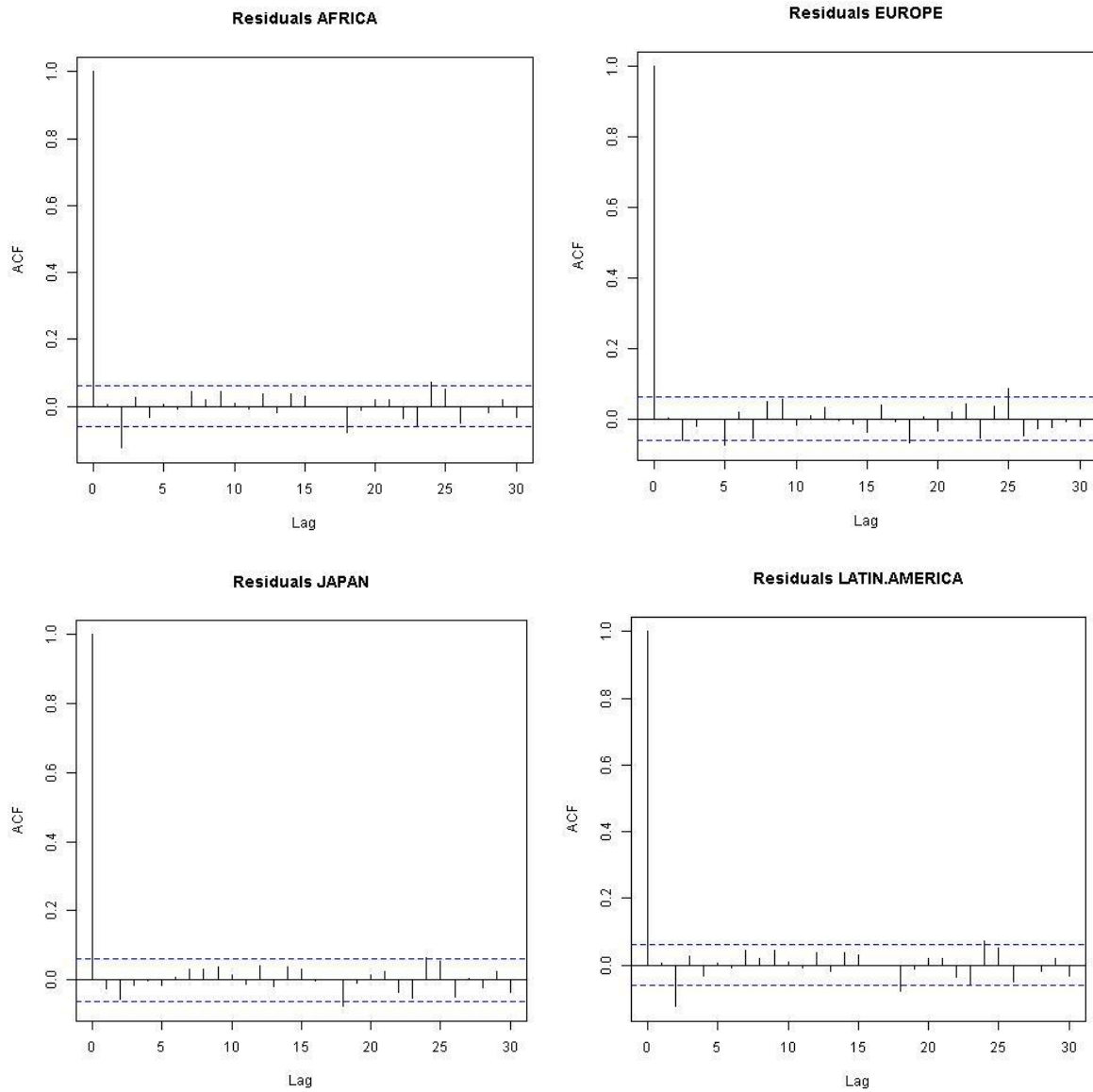
NORTH.AMERICA : Return ACF



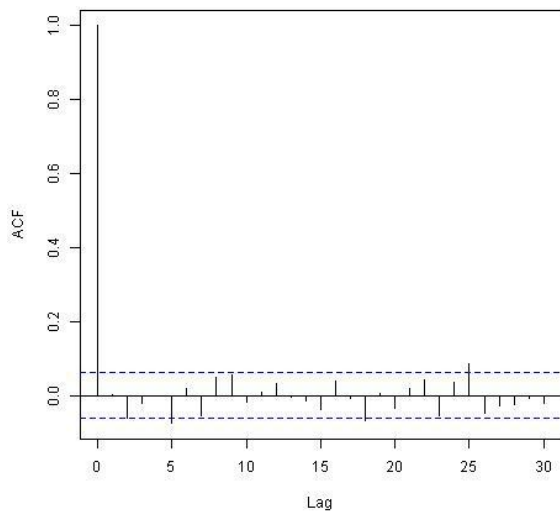
NORTH.AMERICA : Squared Return ACF



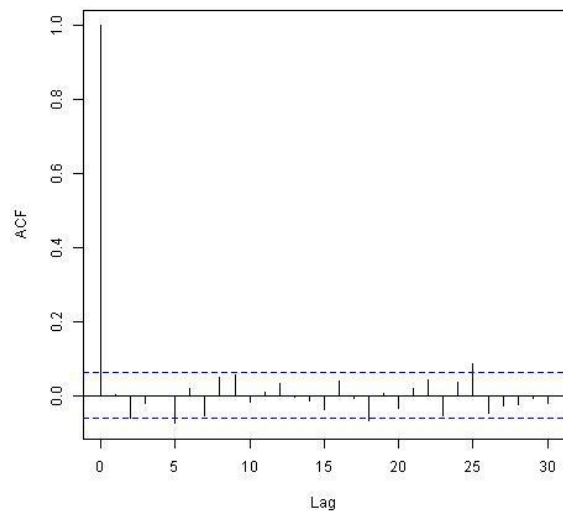
Diagnostic plots of residulas:



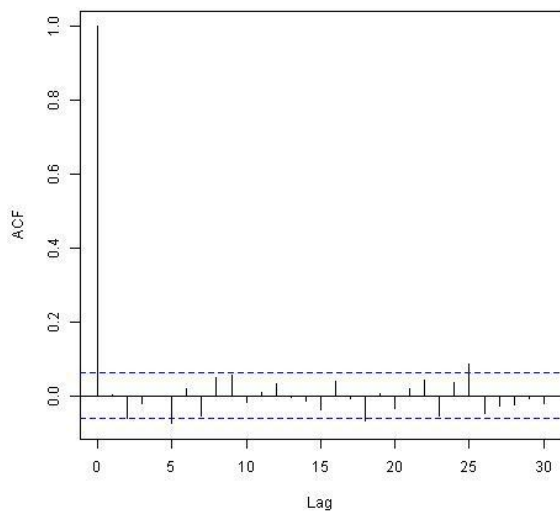
Residuals NORTH.AMERICA



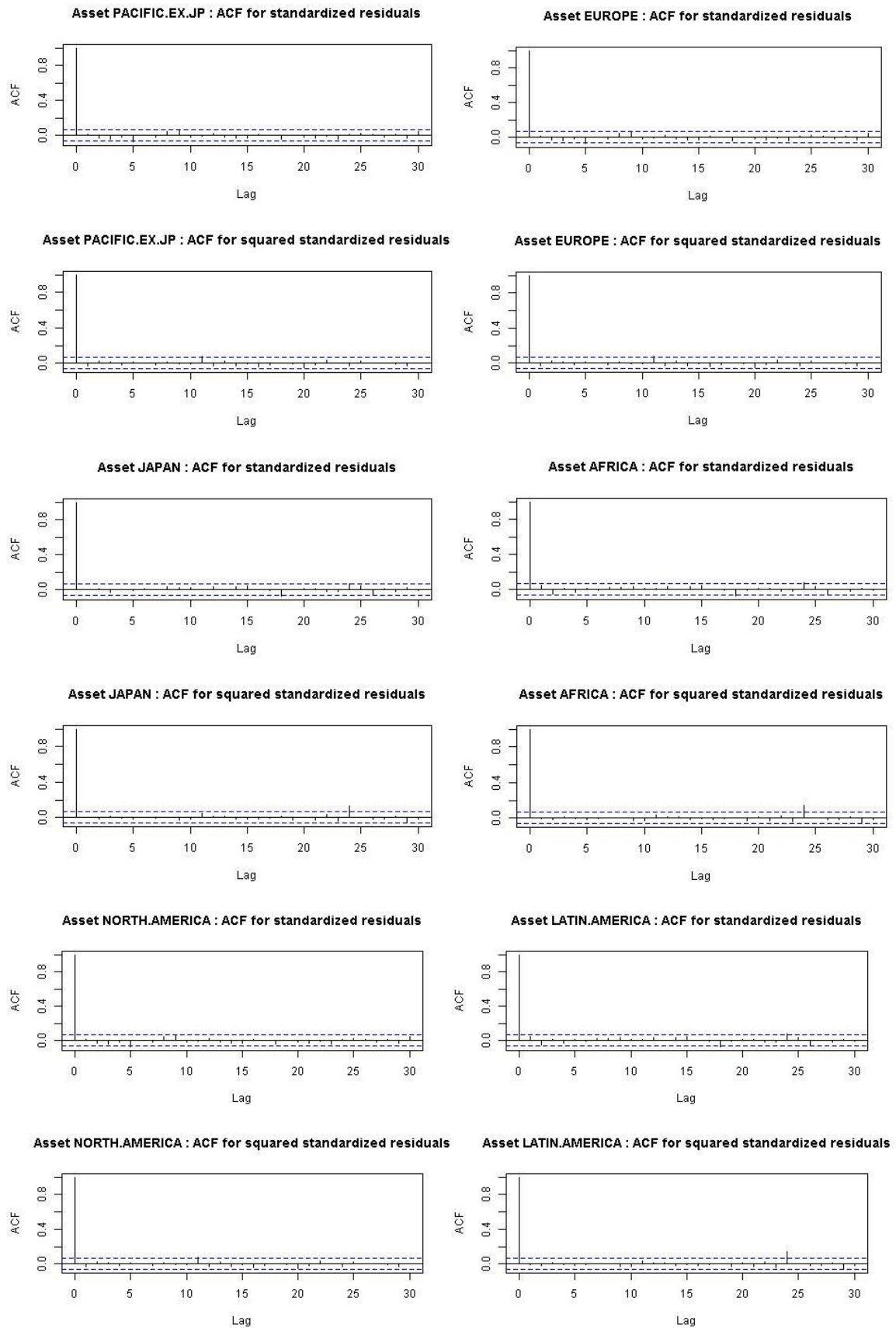
Residuals INDIA



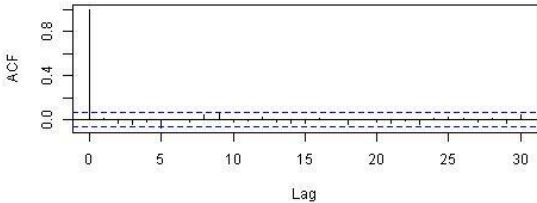
Residuals PACIFIC.EX.JP



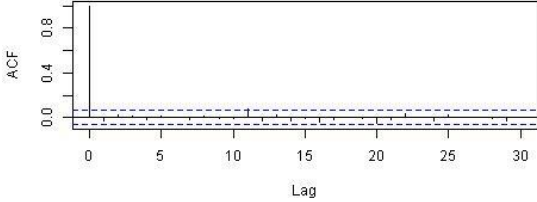
Autocorrelation in univariate Garch plots:



Asset INDIA : ACF for standardized residuals

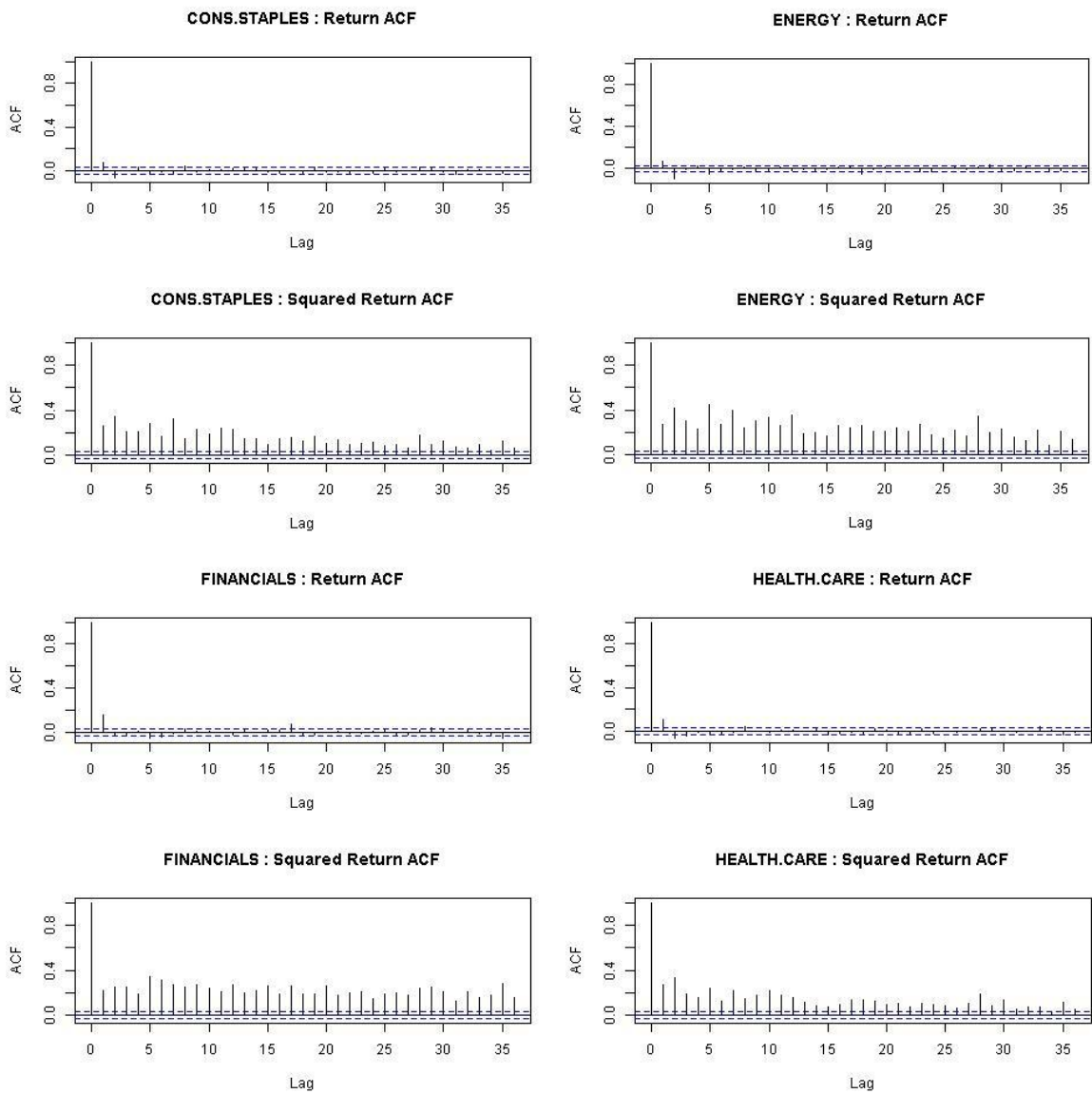


Asset INDIA : ACF for squared standardized residuals

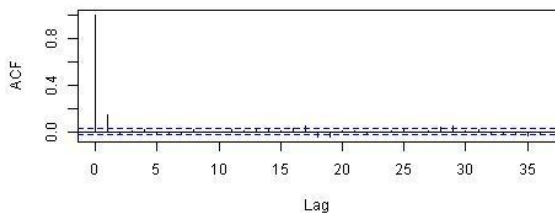


12.3.2 Sector

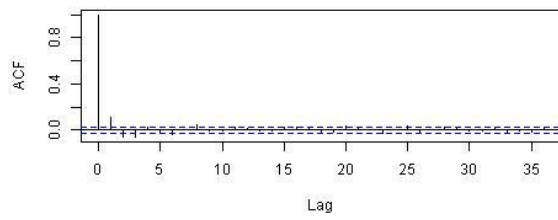
Original series ACF and ACF for squared return:



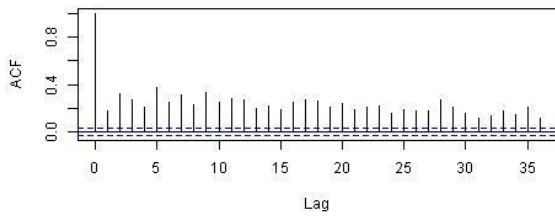
INDUSTRIALS : Return ACF



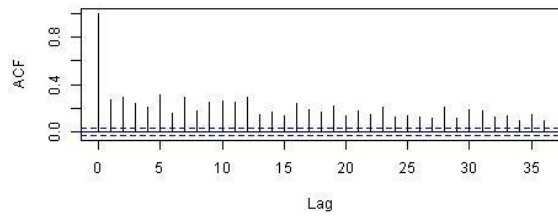
T.CM.SVS : Return ACF



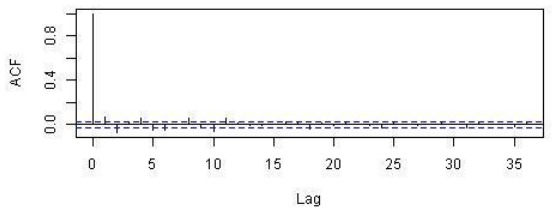
INDUSTRIALS : Squared Return ACF



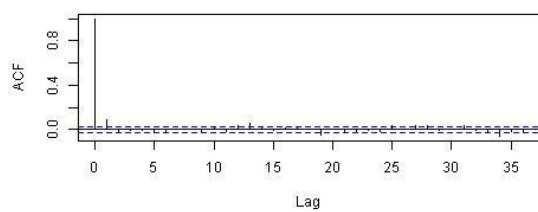
T.CM.SVS : Squared Return ACF



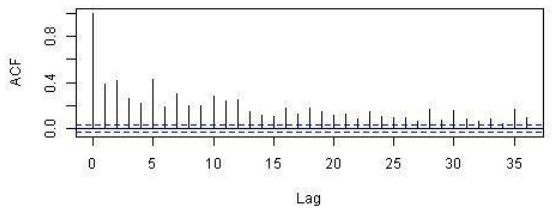
UTILITIES : Return ACF



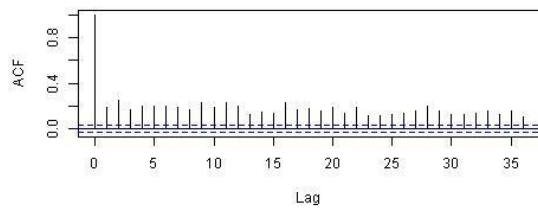
WORLD.IT : Return ACF



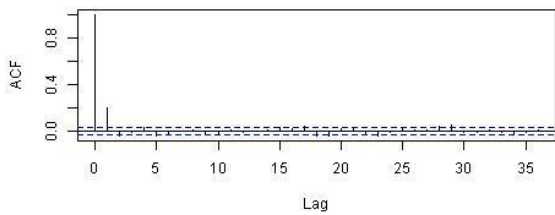
UTILITIES : Squared Return ACF



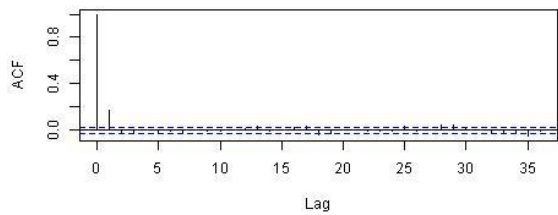
WORLD.IT : Squared Return ACF



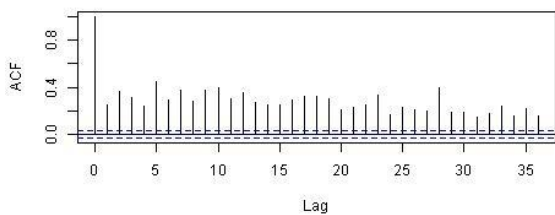
WORLD.MATERIALS : Return ACF



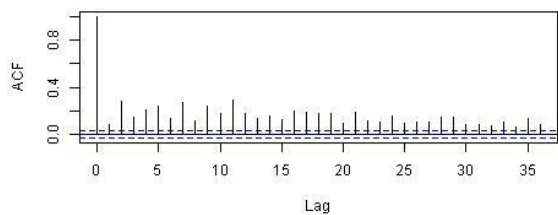
CONS.DISCR : Return ACF



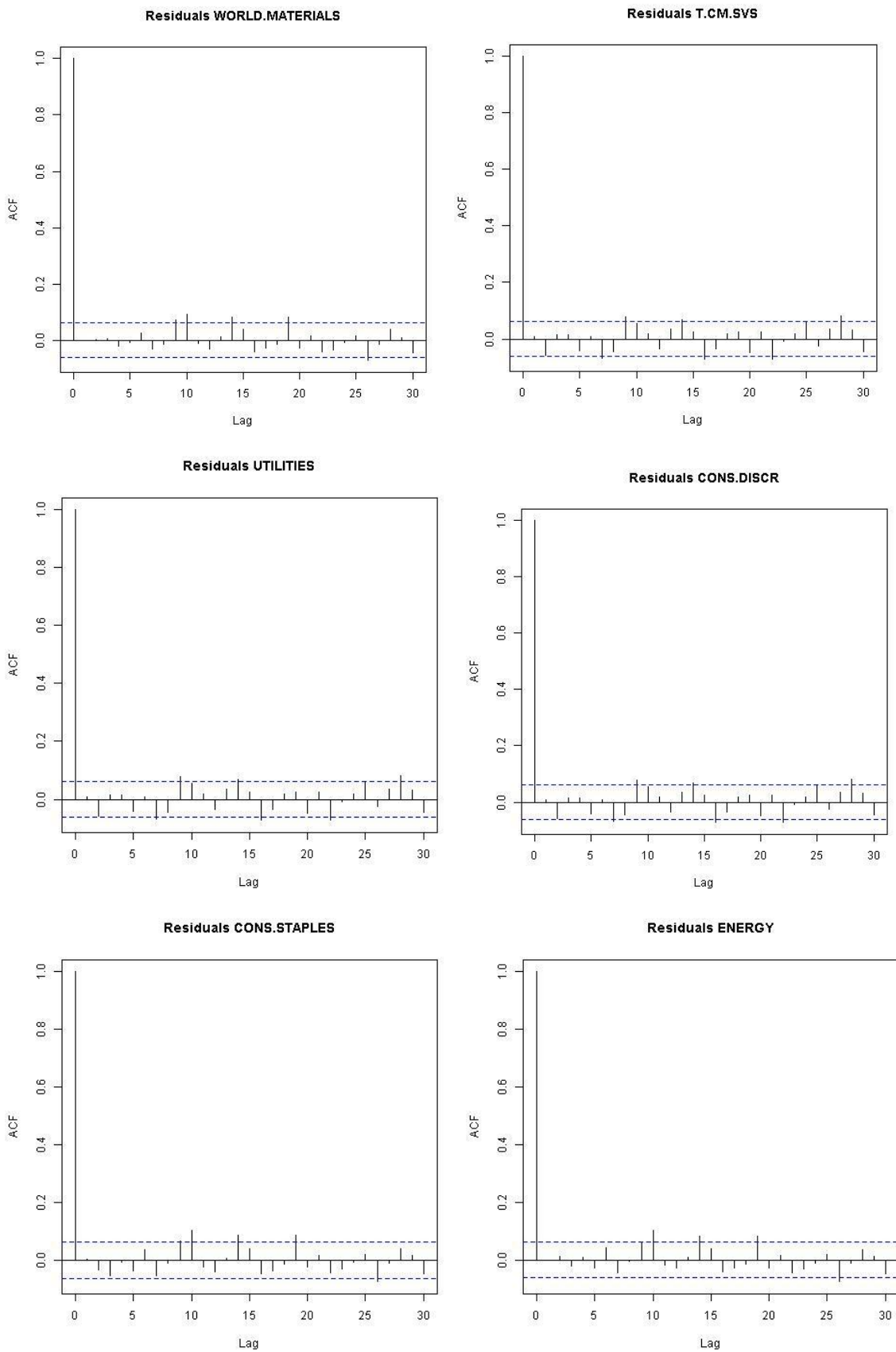
WORLD.MATERIALS : Squared Return ACF



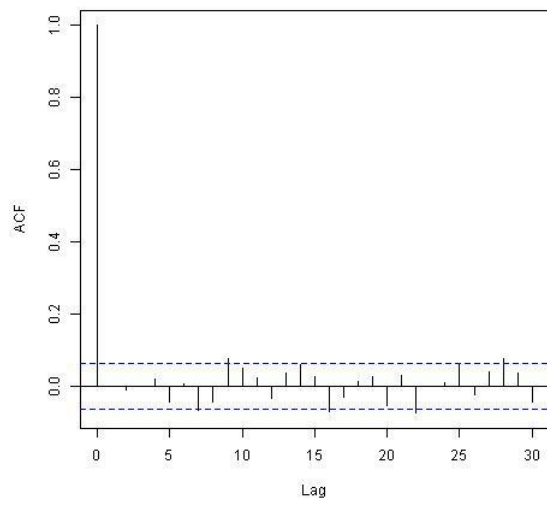
CONS.DISCR : Squared Return ACF



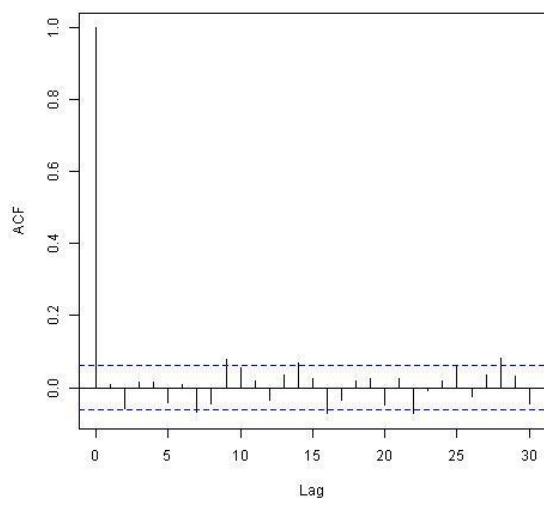
Diagnostic plots of residulas:



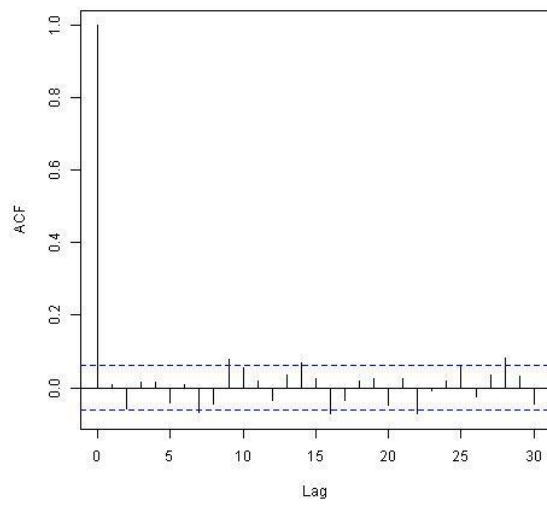
Residuals FINANCIALS



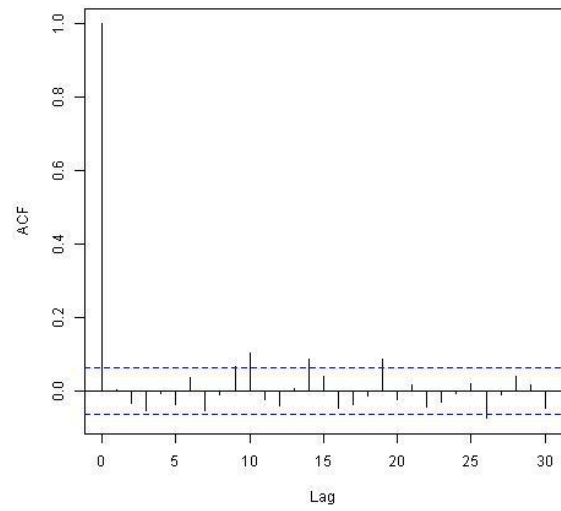
Residuals HEALTH.CARE

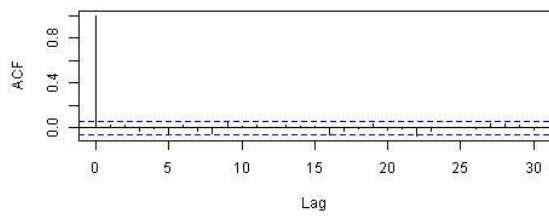
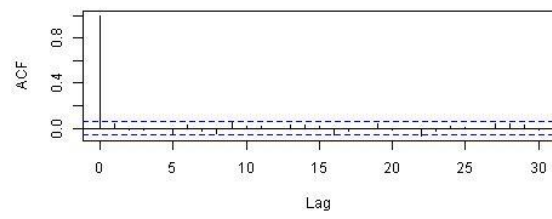
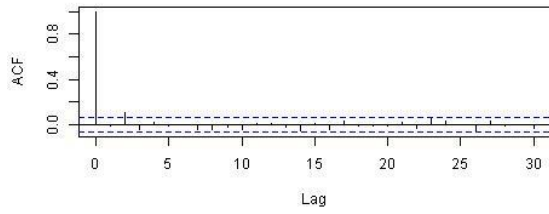
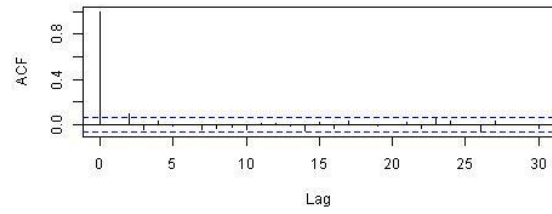
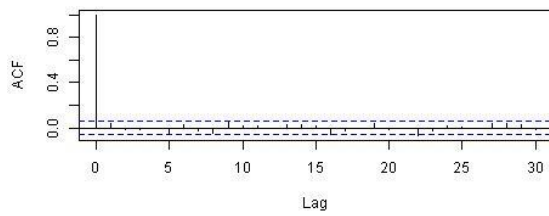
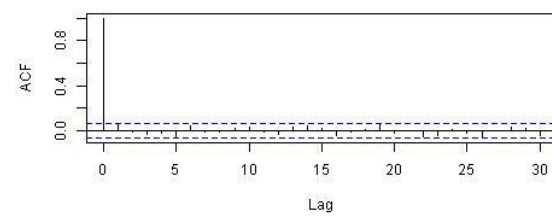
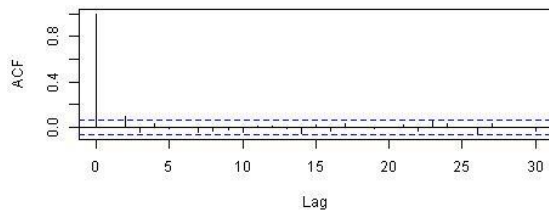
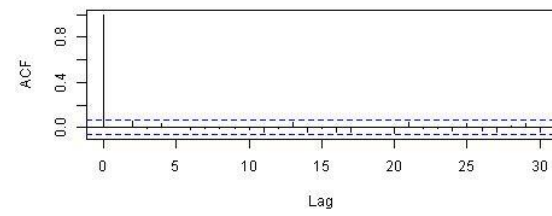


Residuals INDUSTRIALS



Residuals WORLD.IT



Asset FINANCIALS : ACF for standardized residuals**Asset HEALTH.CARE : ACF for standardized residuals****Asset FINANCIALS : ACF for squared standardized residuals****Asset HEALTH.CARE : ACF for squared standardized residuals****Asset INDUSTRIALS : ACF for standardized residuals****Asset WORLD.IT : ACF for standardized residuals****Asset INDUSTRIALS : ACF for squared standardized residuals****Asset WORLD.IT : ACF for squared standardized residuals**

12.4 Appendix D: Deriving VaR

Deriving VaR for continuously compounded returns:

$$Y_t = \log(P_t) - \log(P_{t-1})$$

$$p = \Pr(P_t - P_{t-1} \leq -VaR(p))$$

$$p = \Pr(P_{t-1}(e^{Y_t} - 1) \leq -VaR(p))$$

$$p = \Pr(P_{t-1}(e^{Y_t} - 1) \leq -VaR(p))$$

$$p = \Pr(P_{t-1}e^{Y_t} \leq -VaR(p) + P_{t-1}) \quad | \text{Dividing with } P_{t-1}$$

$$p = \Pr\left(e^{Y_t} \leq -\frac{VaR(p)}{P_{t-1}} + 1\right) \quad | \text{ Taking the logarithm.}$$

$$p = \Pr\left(Y_t \leq \log\left(-\frac{VaR(p)}{P_{t-1}} + 1\right)\right) \quad | \text{dividing with standard deviation on both sides}$$

$$p = \Pr\left(\frac{Y_t}{\sigma} \leq \log\left(-\frac{VaR(p)}{P_{t-1}} + 1\right)\left(\frac{1}{\sigma}\right)\right)$$

Since $-\frac{VaR(p)}{P_{t-1}} \leq 1$. Then the distribution of standardized returns (Y_t/σ) can be denoted

$\Phi_y(\cdot)$ and the inverse distribution by $\gamma(p) = \Phi_y^{-1}(p)$ we have:

$$\log\left(-\frac{VaR(p)}{P_{t-1}} + 1\right) = \Phi^{-1}(p) * \sigma$$

$$VaR(p) = (\exp((\Phi^{-1}(p) * \sigma)) - 1) * P_{t-1}$$

For small $\Phi^{-1}(p) * \sigma$ the VaR is approximately given this way:

$$VaR \approx -\sigma * \Phi^{-1}(p)P_{t-1}$$

Definition of higher moments:

Skewness:

The skewness coefficient is given by:

$$S = \frac{E[r_p^3]}{\sigma^3}$$

The co-skewness is then $N \times N^2$ matrix:

$$m3 = E[(r - \mu)(r - \mu)' \otimes (r - \mu)']$$

\otimes is the kronecker product.

Kurtosis:

$$K = \frac{E[r_p^4]}{\sigma^4}$$

The co kurtosis $N \times N^3$ matrix and is defined the following way:

$$m_4 = E[(r - \mu)(r - \mu)' \otimes (r - \mu)' \otimes (r - \mu)']$$

12.5 Appendix E: Differential Evolution Algorithm

The differential evolution algorithm, introduced by Storn and Price (1997) optimize a problem by maintaining a population of candidate solutions and creating new candidate solutions by combining existing ones, and then keeping the candidate solution which has the best score or fitness on the optimization problem at hand. DE is particularly well suited to find the global optimum of a real-valued function of real-valued parameters, and does not require that the function is neither continuous nor differentiable. There is also easy to add constraints, both linear and non-linear.

The DE algorithm consists of three main steps; mutation, crossover and selection. One set of optimization parameters, called an individual, is represented by a D -dimensional vector. A population consists of NP D -dimensional parameter vectors x_{iG} , $i = 1, 2, 3, \dots, NP$ for each generation G .

Mutation: For each target vector x_{iG} , a mutant vector is generated according to $v_{i,G+1} = x_{r_1,G} + F * (x_{r_2,G} - x_{r_3,G})$ with randomly chosen indexes $r_1, r_2, r_3 \in \{1, 2, 3, \dots, NP\}$. Note that smaller differences between parameters of parent r_2 and r_3 , the smaller the difference vector will be, and therefore the perturbation will be smaller. This means that if the population gets close to optimum, the step length is automatically decreased. The vector generating process can be done in many different ways, called strategies. The strategy we will apply is called the “DE/local-to-best/1/bin” and is specified as follow:

$$v_{i,G+1} = old_{1,g} + F * (best_g - old_{1,g}) + F * (x_{r_2,G} - x_{r_3,G})$$

where $old_{1,g}$ and $best_g$ are the i^{th} member and best member respectively, of the previous population.

Crossover: The target vector is mixed with the mutated vector using the following scheme to yield the trial vector $u_{i,G+1} = (u_{1,i,G+1}, u_{2,i,G+1}, \dots, u_{D,i,G+1})$ where

$$u_{1i,G+1} = \begin{cases} v_{ji,G+1} & \text{if } (r(j) \leq CR) \text{ or } j = rn(i) \\ x_{ji,G} & \text{if } (r(j) > CR) \text{ and } j \neq rn(i) \end{cases}$$

For $j = 1, 2, 3, \dots, D$, $r(j) \in [0, 1]$ is the j^{th} evaluation of a uniform random number generator. CR is the crossover constant $\in [0, 1]$. $CR = 0$ means no crossover. $rn(i) \in (1, 2, \dots, D)$ is a randomly chosen index which ensures that $u_{(i, G+1)}$ gets at least one element from $v_{i, G+1}$. Otherwise no new parent vector would be produced and the population would not alter.

Selection: A greedy selection scheme is used: If and only if the trial vector yields a better cost function value compared to the parameter vector $x_{i, G}$, is it accepted as a new parent vector for the following generation $G + 1$. Otherwise, the target vector is retained to serve as a parent vector for generation $G + 1$ once again.