



UNIVERSITEIT • STELLENBOSCH • UNIVERSITY
jou kennisvennoot • your knowledge partner

VALUE AT RISK AND EXPECTED SHORTFALL:
TRADITIONAL MEASURES AND EXTREME VALUE THEORY
ENHANCEMENTS

WITH A SOUTH AFRICAN MARKET APPLICATION

by

Anelda Dicks

*Assignment presented at the University of Stellenbosch in partial fulfilment of the
requirements for the degree of*

Masters of Commerce

Department of Statistics and Actuarial Science

University of Stellenbosch

Private Bag X1, 7602 Matieland, South Africa

Supervisors: Prof. W.J. Conradie

Prof. T. de Wet

December 2013

Copyright © 2013 University of Stellenbosch

All rights reserved.

Declaration

I, the undersigned, hereby declare that the work contained in this assignment is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

Signature:

A. Dicks

Date:

Abstract

Accurate estimation of Value at Risk (VaR) and Expected Shortfall (ES) is critical in the management of extreme market risks. These risks occur with small probability, but the financial impacts could be large.

Traditional models to estimate VaR and ES are investigated. Following usual practice, 99% 10 day VaR and ES measures are calculated. A comprehensive theoretical background is first provided and then the models are applied to the Africa Financials Index from 29/01/1996 to 30/04/2013. The models considered include independent, identically distributed (i.i.d.) models and Generalized Autoregressive Conditional Heteroscedasticity (GARCH) stochastic volatility models. Extreme Value Theory (EVT) models that focus especially on extreme market returns are also investigated. For this, the Peaks Over Threshold (POT) approach to EVT is followed. For the calculation of VaR, various scaling methods from one day to ten days are considered and their performance evaluated.

The GARCH models fail to converge during periods of extreme returns. During these periods, EVT forecast results may be used. As a novel approach, this study considers the augmentation of the GARCH models with EVT forecasts. The two-step procedure of pre-filtering with a GARCH model and then applying EVT, as suggested by McNeil (1999), is also investigated.

This study identifies some of the practical issues in model fitting. It is shown that no single forecasting model is universally optimal and the choice will depend on the nature of the data. For this data series, the best approach was to augment the GARCH stochastic volatility models with EVT forecasts during periods where the first do not converge. Model performance is judged by the actual number of VaR and ES violations compared to the expected number. The expected number is taken as the number of return observations over the entire sample period, multiplied by 0.01 for 99% VaR and ES calculations.

Opsomming

Akkurate beraming van Waarde op Risiko (*Value at Risk*) en Verwagte Tekort (*Expected Shortfall*) is krities vir die bestuur van ekstreme mark risiko's. Hierdie risiko's kom met klein waarskynlikheid voor, maar die finansiële impakte is potensieel groot.

Tradisionele modelle om Waarde op Risiko en Verwagte Tekort te beraam, word ondersoek. In ooreenstemming met die algemene praktyk, word 99% 10 dag maatstawwe bereken. 'n Omvattende teoretiese agtergrond word eers gegee en daarna word die modelle toegepas op die *Africa Financials Index* vanaf 29/01/1996 tot 30/04/2013. Die modelle wat oorweeg word sluit onafhanklike, identies verdeelde modelle en Veralgemeende Auto-regressiewe Voorwaardelike Heteroskedastiese (*GARCH*) stogastiese volatiliteitsmodelle in. Ekstreemwaarde Teorie modelle, wat spesifiek op ekstreme mark opbrengste fokus, word ook ondersoek. In hierdie verband word die *Peaks Over Threshold (POT)* benadering tot Ekstreemwaarde Teorie gevolg. Vir die berekening van Waarde op Risiko word verskillende skaleringsmetodes van een dag na tien dae oorweeg en die prestasie van elk word ge-evalueer.

Die *GARCH* modelle konvergeer nie gedurende tydperke van ekstreme opbrengste nie. Gedurende hierdie tydperke, kan Ekstreemwaarde Teorie modelle gebruik word. As 'n nuwe benadering oorweeg hierdie studie die aanvulling van die *GARCH* modelle met Ekstreemwaarde Teorie vooruitskattings. Die sogenaamde twee-stap prosedure wat voor-af filtrering met 'n *GARCH* model behels, gevolg deur die toepassing van Ekstreemwaarde Teorie (soos voorgestel deur McNeil, 1999), word ook ondersoek.

Hierdie studie identifiseer sommige van die praktiese probleme in model passing. Daar word gewys dat geen enkele vooruiskattingsmodel universeel optimaal is nie en die keuse van die model hang af van die aard van die data. Die beste benadering vir die data reeks wat in hierdie studie gebruik word, was om die *GARCH* stogastiese volatiliteitsmodelle met Ekstreemwaarde Teorie vooruitskattings aan te vul waar die voorafgenoemde nie konvergeer nie. Die prestasie van die modelle word beoordeel deur die werklike aantal Waarde op Risiko en Verwagte Tekort oortredings met die verwagte aantal te vergelyk. Die verwagte aantal word geneem as die aantal obrengste waargeneem oor die hele steekproefperiode, vermenigvuldig met 0.01 vir die 99% Waarde op Risiko en Verwagte Tekort berekeninge.

Acknowledgements

The author thanks the following people for their contribution towards this study:

- Prof. Tertuis de Wet, my co-study leader for theoretical assistance and proof reading as well as for moral support and encouragement throughout the research period. Thank you for unforgettable study years and valuable life insights.
- Prof. Willie Conradie, co-study leader for research suggestions, meticulous proof reading and guidance during my research. Your passion about financial risk management is inspiring. Thank you.
- Prof. N.J. le Roux for advice on plots in R.
- My family for their patience and emotional support.
- My friends for all their interest and for making this journey a pleasant one.

Contents

Declaration	iii
Abstract	iv
Opsomming	v
Acknowledgements	vi
Contents	vii
List of Abbreviations	xii
List of Notation	xiii
1. Introduction	1
2. Brief Literature Review of Extreme Value Theory Estimation Approaches	3
2.1. General results	3
2.2. Alternatives to the GPD	5
2.3. Empirical quantile approach	6
2.4. Semi-nonparametric EVT approach	6

2.5. Alternative to dynamic back-testing	7
3. Traditional VaR and ES measures	8
3.1. Basic Definitions	8
3.2. Comparison of VaR and ES	10
3.3. Models for independent, identically distributed returns	11
3.3.1. Normal model	11
3.3.2. Student t model	13
3.3.3. The problem with the i.i.d. assumption	13
3.4. The GARCH conditional model	14
3.4.1. Model description	14
3.4.2. Maximum likelihood estimation	16
3.4.2.1. Normal model	17
3.4.2.2. Student t model	18
3.4.3. Forecasts and long term volatility	20
3.4.3.1. Long term volatility	21
3.4.3.2. One day ahead forecasts	21
3.4.3.3. Term structure of volatility forecasts	23
3.4.3.4. Convergence of forecasts to long term volatility	23
3.5. Asymmetric GARCH conditional models	24
3.5.1. The rationale for asymmetric models	24
3.5.2. A-GARCH and GJR-GARCH	25
3.5.2.1. Model definition, long term variance and forecasts	25
3.5.2.2. Estimation	26

3.5.3. E-GARCH	26
3.5.3.1. Model definition	26
3.5.3.2. Properties	27
3.5.3.3. Long term variance	28
3.5.3.4. Estimation	29
3.5.3.5. One day forecasts	30
3.6. Summary	32
4 Traditional EVT	33
4.1 Introduction	33
4.2 Method of Block Maxima	34
4.3 Peaks over Threshold (POT) approach	37
4.3.1 Theoretical motivation for using the GPD	38
4.3.2 Generalised Pareto Distribution (GPD)	39
4.3.2.1 Parameter estimation	39
4.4 Fréchet class of distributions	42
4.4.1 Basic Definitions	43
4.4.2 The extreme value index	44
4.4.3 Threshold selection and exploratory data analysis	46
4.4.3.1. Quantile-Quantile (QQ) plots	47
4.4.3.2. Hill plots	48
4.4.3.3. Mean excess plots	48
4.5 Quantile estimation	49
4.5.1 First order estimation	50

4.5.1.1	The Weissman estimator	50
4.5.1.2	The probability approach estimator	52
4.5.2	Second order refinements	54
4.5.2.1	Quantile view	54
4.5.2.2	Probability view	57
4.6	Derivation of VaR and ES using the POT approach	60
4.6.1	Confidence intervals for VaR and ES	62
4.7	Limitations of EVT	65
4.8	Scaling of VaR and ES measures	66
4.8.1	The square root of time rule	66
4.8.2	The alpha root rule	66
4.8.3	Empirically established scaling laws	67
4.9	Summary	68
5.	Combining EVT with stochastic volatility models	70
5.1.	Two step procedure	70
5.3.	h-day time horizons	72
6.	Practical issues in model fitting	73
7.	The data	75
7.1	Description of the data	75

8. Analysis of the Africa Financials Index Data	81
8.1. Schematic Overview	81
8.2. Independent identically distributed models	82
8.3. GARCH models	85
8.3.1. Estimation results	85
8.3.2. Analysis of the absolute exceedances	89
8.4. Extreme Value Analysis	93
8.4.1. Exploratory analysis for the first data window	93
8.4.2. VaR and ES over the whole data series	105
8.4.2.1. One day VaR and ES estimates	107
8.4.2.2. Ten day VaR and ES estimates	109
8.4.3. Augmenting GARCH forecasts with EVT forecasts	116
8.5. Two step procedure	125
8.5.1. Filter with the i.i.d. normal model	126
8.5.2. Filter with the i.i.d. Student t model	127
8.5.3. Filter with the symmetric GARCH Student t model	130
9. Conclusion and Open Questions	135
Appendix: R Code for Chapters 7 and 8	140
Bibliography	216

List of Abbreviations

ACF	Auto Correlation Function
A-GARCH	Asymmetric GARCH
AR	Auto Regressive
ARMA	Auto Regressive Moving Average
c.d.f.	Cumulative distribution function
E-GARCH	Exponential GARCH
ES	Expected Shortfall
EVI	Extreme Value Index
EVT	Extreme Value Theory
GARCH	Generalized Autoregressive Conditional Heteroscedasticity
GEV	Generalised Extreme Value
GJR-GARCH	Glosten Jagannathan and Runkle GARCH
GPD	Generalised Pareto Distribution
HTE	Heavier Tail than Exponential
i.i.d.	Independent identically distributed
LTE	Lighter Tail than Exponential
MA	Moving Average
ML	Maximum Likelihood
MOM	Method Of Moments
p.d.f.	Probability density function
POT	Peaks Over Threshold
PPD	Perturbed Pareto Distribution
PWM	Probability Weighted Moments
QQ	Quantile-Quantile
VaR	Value at Risk

List of Notation

The following summarises the most important notation used in this paper.

Chapter 3

R_t	The backward-looking daily log-return of a portfolio
	$R_t = (\ln P_t - \ln P_{(t-1)}) \approx \frac{(P_t - P_{(t-1)})}{P_{(t-1)}}$
P_t	The portfolio price at time t .
$x_{t,\alpha}$	The α^{th} percentile of the distribution of the return of the portfolio at time t .
VaR_α	The $100(1-\alpha)\%$ daily Value at Risk of a portfolio.
ES_α	The $100(1-\alpha)\%$ daily Expected Shortfall of a portfolio.
$f_R(r)$	The probability density function of the return distribution.
$\Phi(\cdot)$	The standard normal cumulative distribution function.
$\varphi(\cdot)$	The standard normal probability density function.
μ_t	The expected value of R_t .
σ_t^2	The variance of R_t .
Z_t	A random variable having either a standard normal or standardised Student t distribution.
E_t	The random error term in the model for R_t .
ε_t	The observed value of E_t .

φ_0, φ_1	The AR(1) parameters in the conditional mean equation for R_t .
r_t	The observed value of R_t .
I_{t-1}	The information set of all past returns up to and including time $t-1$.
ω, α, β	Constant parameter, error parameter and lag parameter in the GARCH(1,1) conditional variance equation for σ_t^2 .
$\hat{\omega}, \hat{\alpha}, \hat{\beta}$	The corresponding maximum likelihood estimates.
$f_v(t)$	Standardised student t distribution function.
$\bar{\sigma}^2$	Long-term volatility of GARCH models.
λ	Leverage parameter in Asymmetric GARCH models.

Chapter 4

$X_{i,n}$	The i^{th} order statistic of a sample of size n .
$F(x)$	Distribution function $F(x) = P(X \leq x)$
$Q(p)$	The quantile function of a random variable $X \sim F$ $Q(p) := \inf\{x : F(x) \geq p\}.$
$U(x)$	The tail quantile function $U(x) = Q\left(1 - \frac{1}{x}\right).$
X_{nn}	The maximum of a sample of size n .
$G_{\sigma,\gamma,\mu}(x), G_\gamma$	The Generalised Extreme Value distribution.
γ	The Extreme Value Index.

\hat{R}_n^k	The estimated return level.
u	The threshold value.
Y_j	The absolute exceedances above the threshold value $\{Y_j = X_i - u \text{ given } X_i > u\}_{j=1}^{N_u}$
N_u, k	The number of exceedances.
$D(G_\gamma)$	The domain of attraction of G_γ .
$H_{\gamma,\sigma}(y), h_{\gamma,\sigma}(y)$	Generalised Pareto Distribution function and density function.
$\hat{\gamma}_{MOM}, \hat{\sigma}_{MOM}$	Method of moments estimators for the parameters of the GPD.
$M_{p,r,s}$	The $(p, r, s)^{\text{th}}$ probability weighted moment.
$\hat{\gamma}_{PWM}, \hat{\sigma}_{PWM}$	Method of probability weighted moments estimators for the parameters of the GPD.
$A(\cdot)$	A function of regular variation.
$\ell(\cdot)$	A slowly varying function.
$\hat{\gamma}_{Hill} = H_{k,n}$	The Hill estimator of the Extreme Value Index.
$M_{k,n}$	The moment estimator of the Extreme Value Index.
$e(\cdot)$	The mean excess function $e(u) = E(X - u X > u) = E(X X > u) - u$
$\hat{q}_{k,p}^+$	The Weissman (1978) estimator for extreme quantiles.
H_0	Null hypothesis.
χ_1^2	Chi-squared distribution with one degree of freedom.
$\chi_1^2(\alpha)$	The α^{th} percentile of a Chi-squared distribution with one degree of

freedom.

X_t Return level for EVT models

$$X_t = -R_t = (\ln P_{t-1} - \ln P_t) \approx \frac{P_{t-1} - P_t}{P_{t-1}}.$$

F_X The distribution function of the daily percentage returns X_t .

n The total number of returns.

F_U The conditional distribution of the absolute exceedances.

Chapter 5

X_i The return at time i in the model for the two step procedure.

μ_i, σ_i The mean and standard deviation of X_i in the two step procedure.

Z_i The residual from the two step procedure.

1. Introduction

The modelling of extreme market risks and the management of the impact of it are important topics in financial risk management. Extreme market risk is risk due to extreme changes in prices (Ruppert, 2004), e.g. stock market crashes. These risks occur with small probability, but have large financial consequences. In order to study and understand the risk with respect to the extreme market events, the estimation of the daily Value at Risk (VaR) and Expected Shortfall (ES) measures are of special interest. To estimate these quantities, extreme risks need to be modelled. The model assumed for the volatility of the returns is also of particular importance. Various models for the variance or extreme risks are investigated in this thesis.

The determination of VaR or ES is only one of the fields in which the management of extreme risks is useful. Other fields are credit or operational risk management and insurance risks (McNeil, 1999). In credit risk management the size of reserves to be held to guard against unexpected credit events need to be determined (e.g. credit downgradings). Reserves for excess-of-loss reinsurance products also deal with extreme risks (in this case catastrophic losses).

The aim of this thesis is to provide a detailed theoretical overview of both traditional Value at Risk models and Extreme Value Theory. For this purpose, the Africa Financials Index from 29/01/1996 to 30/04/2013 is considered. Since the practical implementation of these models is very important, the thesis aims to identify what works in practice as well as to highlight the short comings of the various models. A combination of forecasting techniques is suggested to augment areas where traditional methods fail. For example, Extreme Value Theory may sensibly improve Value at Risk forecasts during periods of extreme returns. The so-called two-step procedure suggested by McNeil (1999) is also investigated and the performance of this procedure on the Africa Financial Index is evaluated. Expected Shortfall estimates are similarly considered.

The models considered to produce VaR and ES estimates fall into five classes. The traditional normal model and the traditional Student t model constitute the first class. These models assume independent, identically distributed returns and yield constant one-day estimates. The second class is stochastic volatility models. The mean and variance equations of the returns are allowed to be time varying and these are modelled by AR and

GARCH models. The innovations are assumed to be either normally, or Student t distributed. The symmetric GARCH, GJR-GARCH and E-GARCH are considered. The third class of models is based on Extreme Value Theory. A Generalised Pareto distribution (GPD) is fitted to the returns and estimated GPD percentiles are used in the estimation of VaR and ES. Different scaling methods of one day to ten 10 estimates are investigated and compared. The fourth class of models augments the GARCH models with EVT forecasts during the periods where the GARCH models do not converge. The fifth model is a hybrid between the stochastic volatility model and the Extreme Value GPD model. A GPD is fitted after first filtering the returns by fitting AR and GARCH conditional mean and variance equations. Throughout, the estimation procedure is dynamic, using a 250 days moving data window. Dynamic back-testing is used to evaluate the performance of the different models.

The remainder of the thesis is structured as follows. In chapter two a literature overview is provided. In chapter three the traditional Value at Risk and Expected shortfall measures are presented, and in chapter four Extreme Value Theory is considered. In chapter five the combination of Extreme Value Theory with stochastic volatility models is discussed. In chapter six practical issues in model fitting are highlighted. In chapter seven the data that will be used for the South African market example is described. Throughout this thesis, the Africa Financials Index data from 29/01/1996 to 30/04/2013 is considered. The data analysis follows in chapter eight. In chapter nine a summary and discussion of the results is presented. Finally, chapter ten contains concluding remarks and notes regarding further work.

2. Brief Literature Review of Extreme Value Theory Estimation Approaches

2.1. General results

McNeil (1999) performed ground-breaking work in the combination of EVT with stochastic volatility models. He introduced the so-called two-stage procedure for estimating VaR and ES. (Refer chapter 5 for detail of this approach). The first step is to filter the return observations by fitting a GARCH type stochastic volatility model using maximum likelihood. The second step is to apply EVT to the residuals and to model these residuals with a GPD (Generalised Pareto distribution). The GPD quantiles can then be used in conjunction with the dynamic estimates of the mean and variance of returns to obtain VaR and ES estimates.

McNeil (1999) studied the daily losses of the DAX and S&P indices. He suggests a dynamic back-testing period of 1 000 days and a threshold (to fit the GPD) equal to the 90th sample percentile of the residuals. In his study he found that dynamic (conditional) EVT yielded the best estimates, especially for high significance levels. The traditional Student t VaR performed reasonably well, provided that the return distribution was not overly asymmetric in nature. For significance levels higher than 99%, the Normal VaR performed very poorly. McNeil does not recommend using the Normal ES estimate, even for lower significance levels. The only sensible ES estimates were those that used EVT.

McNeil noted that the static (unconditional) EVT estimate of VaR changes markedly slower than the dynamic (conditional) EVT estimate. This is because an extreme observation will remain for n days in the n -day data window and affect the estimate. The static EVT estimate will therefore be violated several times in a row, if a violation occurs. The dynamic EVT estimate of VaR changes quickly.

Fernandez (2003) calculated VaR estimates for four time series of the Chilean financial market. The Index Price of Selective stocks (1990-2002), the Chilean peso/US-dollar exchange rate (1988-2002), the spot price of copper (1988-2002) and one year domestic zero-coupon bonds (1993-2001) were considered. Significance levels were taken to be 95, 99 and 99.5% for all series.

Conditional EVT was found to produce the least number of significant violations. (A violation is said to occur if the observed loss exceeds the estimated loss. Significance is measured by a binomial test statistic. Refer chapter 5.) The traditional Student t VaR also produced good results. The traditional Normal VaR performed by far the poorest, and this poor performance was exacerbated at higher significance levels. However, further studies by Fernandez revealed that the t -VaR cannot replace the conditional EVT VaR estimate.

It is interesting to note that the study by Fernandez (2003) found that conditional EVT measures performed poorer for series characterised by low volatility and high kurtosis. This is possibly because there are fewer outliers in such series and the tail estimation is less effective.

The unconditional EVT approach may be better for longer time horizons (Gilli *et al.*, 2006) or the study of stress events (very rare events) (Byström, 2004) as less frequent updating is required and stable estimates preferred. However, for shorter time horizons (a few hours or days), conditional EVT is recommended.

The study by Byström (2004) of the AFF (Swedish Affärvärlens General Index) and DOW (U.S. Dow Jones Industrial Average) from 1980 to 1999 included two high volatility periods. The results showed an underestimation of risk for all the models. For the 95% VaR estimates, the EVT models (conditional and unconditional) and the traditional Normal VaR or t -VaR models performed equally well. However, for the 99% estimates, the EVT models again out-performed the other methods. What is interesting, is that Byström found no significant improvement of the conditional EVT model over the unconditional EVT model. The AR-GARCH model with t innovations also provided reliable results, although it was found to overestimate VaR at lower confidence levels. The Normal AR-GARCH model, on the other hand, overestimated at high confidence levels.

Byström (2004) remarks that conditional estimates (GARCH volatility estimates and conditional EVT estimates) are especially useful in volatile periods. This is due to their ability to change rapidly and to capture the volatility pattern. A large percentage of the extreme losses were predicted by the conditional models. Unconditional models (Normal and Student t models as well as unconditional EVT models) tend to underestimate risks during volatile periods, but overestimate risks during tranquil periods (Byström, 2004).

2.2. Alternatives to the GPD

As an alternative to the GPD (Generalised Pareto Distribution) which results from the so-called Peaks-over-Threshold (POT) approach to EVT, the literature also considers the Block Maxima approach. Briefly, the method consists of dividing the sample data into n blocks of equal length (the block length being a critical consideration) and then determining the maximum value of each block. This leads to the fitting of a Generalised Extreme Value (GEV) distribution.

Although it is the general consensus that the Block Maxima approach does not make optimal use of the data, since it considers only the maximum observation in each block and ignores the rest, McNeil (1999) points out its use in the estimation of return levels (stress losses). The GEV distribution is fitted to a series of maxima, typically by maximum likelihood methods, and the return level is defined as a quantile of the GEV distribution (refer section 4.2 and definition 4.2.3).

The Block Maxima conditional EVT approach (i.e. applying the two step procedure where the second step fits a GEV to the residuals) was investigated by Byström (2004), but dynamic back-testing results did not indicate a significant difference between the POT method and the Block Maxima method.

Furthermore, extreme quantiles can also be estimated using the extreme value index. This leads to a broad field, since there are numerous estimators of the extreme value index, the most popular of these being the Hill estimate (Hill, 1975). The study by Brooks *et al.* (2003) showed that these procedures tend to over-estimate VaR for long futures positions and under-estimate VaR for short positions.

2.3. Empirical quantile approach

Another model that can be used to compute VaR and ES, is the empirical quantile approach (Fernandez, 2003). Residuals are calculated by first fitting AR and GARCH models to the original return series. However, instead of parameterising the tails of the residual distribution by an EVT distribution, empirical quantiles are computed from the empirical distribution of the residuals. These are then used together with the AR and GARCH estimates of the mean and variance to obtain the VaR and the ES estimates.

The advantages of this approach are ease of computation and that it takes account of the thick tails of the return distribution. Fernandez (2003) concluded that the empirical quantile approach is reasonable to use in most cases and performs better than the Normal VaR measure.

A similar approach used by Brooks *et al.* (2003) on futures contract data, revealed that VaRs calculated from empirical quantiles after fitting GARCH models tend to be higher than other estimates based on standard nonparametric extreme value tail estimation approaches, due to the persistence in volatility of a GARCH model. A large return causes the estimate to remain elevated for a longer period.

2.4. Semi-nonparametric EVT approach (Brooks *et al.*, 2003)

This approach consists of fitting two different GPDs to both the upper tail and the lower tail of the log-return distribution. Maximum likelihood estimation is used. The upper and lower VaR measures (say VaR_U and VaR_L) can then be derived as the quantiles from these two GPDs (refer section 4.6 and result 4.6.1).

The upper and lower threshold levels are chosen as follows. Calculate the standard deviation (say $\hat{\sigma}$) of the log-return distribution. Next, use a normality approach and set:

$$\text{Upper threshold} = x_u = 1.96 * \hat{\sigma};$$

$$\text{Lower threshold} = x_L = -1.96 * \hat{\sigma}.$$

Thus 5% of the data is captured in the lower tail and 5% in the upper tail. This leaves 90% of the data in the centre of the distribution and this is modelled by an empirical distribution. The

reason why the GPD is used for the tails, rather than the empirical distribution, is that the data in the tails are sparse.

A simulation study is performed by bootstrapping from the two GPDs and from the empirical distribution. The details can be found in the paper by Brooks *et al.* (2003). The broad idea is as follows:

1. Draw x_t from the empirical distribution of the log-returns.
2. If $x_t < VaR_L$, then draw from the GPD fitted to the lower tail; if $x_t > VaR_U$, then draw from the GPD fitted to the upper tail; else, the value of x_t is retained.
3. Repeat steps 1 and 2 a large number of times to obtain a simulated distribution of losses. Record the maximum loss (lowest return).
4. Repeating steps 1 to 3 to obtain a distribution of the maximum loss. This distribution is then mapped to a standard normal distribution e.g. by matching the moments to a Johnson system of distributions (for more detail, refer Brooks *et al.*, 2003).

This approach was used by Brooks *et al.* (2003) to calculate VaRs for three futures contracts traded on LIFFE for the period 24 May 1991 to 3 September 1997. Daily log-returns were used. The results from this study show that the semi-nonparametric EVT approach provides superior results compared to other approaches considered. These included the empirical quantile approach and three approaches based on different estimators of the extreme value index.

2.3. Alternative to dynamic back-testing

Instead of dynamic back-testing, it is also possible to evaluate the performance of different models by dividing the sample data period into a training period and a back-test period. The training period returns are used for model estimation. These estimates are then held constant and the returns from the back-test period are used for validation. Fernandez (2003) explored this approach, but Engle (2001) remarks that this does not allow new information from the back-testing period to be incorporated into parameter estimates and quantiles. For this reason, dynamic back-testing (using a moving n day data window) is preferred.

3. Traditional VaR and ES measures

In this chapter some models traditionally assumed for Value at Risk (VaR) and Expected Shortfall (ES) are presented. The simplest of these are models that assume independent and identically distributed returns. However, this assumption is often unrealistic in practice and therefore conditional volatility models are often used. Conditional volatility models model the mean and variance of returns as functions of time. The AR-GARCH model is discussed, as well as various asymmetric conditional volatility models.

3.1 Basic Definitions

In this section the definitions of the portfolio return, VaR and ES as it will be used in this chapter are given and explained.

Definition 3.1.1: Portfolio return

Let P_t be the value of the portfolio at time t , then the backward-looking daily log-return of the portfolio, R_t , is defined as

$$R_t = (\ln P_t - \ln P_{t-1}) \approx \frac{(P_t - P_{t-1})}{P_{t-1}} \quad (3.1) \quad \blacksquare$$

Note that discounting over daily periods is ignored.

Definition 3.1.2: Value at Risk

The $100(1-\alpha)\%$ daily Value at Risk of a portfolio, VaR, is defined as the negative of the α^{th} percentile of the distribution of the return of the portfolio. That is,

$$VaR_\alpha = -x_{t,\alpha}$$

where

$$P(R_t < x_{t,\alpha}) = \alpha \quad (3.2)$$

$$\Rightarrow P(-R_t > VaR_\alpha) = \alpha \quad \blacksquare$$

Alpha, α , is typically chosen to be small, so that $x_{t,\alpha}$ is a percentile situated in the far left of the return distribution and has a negative value. The probability that the return will not fall

below $-|x_{t,\alpha}|$ is $100(1-\alpha)\%$, or equivalently, the probability that the percentage loss (a positive number) in the portfolio will not be greater than VaR_α is $100(1-\alpha)\%$. Defining VaR as a positive value leads to the convention of taking $VaR = -x_{t,\alpha}$.

The Basle Committee requires a 99% VaR over a 10 day time horizon, implying that the distribution of the 10 day return is needed.

VaR is a popular risk measure that asks the question “How bad can things get?”. However, it is often more of interest to know “If things do get bad, how much can the company expect to lose?” (Hull, 2012). Expected Shortfall (ES) answers this last question.

The Expected Shortfall is the expected percentage loss (a positive number) of the portfolio, given that the loss is greater than the VaR percentage (also a positive number). This is therefore the negative (in order to obtain a positive value) of the expected value of the distribution of the return, given that returns are in the lower α quantile of the return distribution. In terms of symbols the Expected Shortfall is defined as:

Definition 3.1.3: Expected Shortfall

$$\begin{aligned} ES_\alpha &= -E\left(R_t \mid R_t < x_{t,\alpha}\right) & (3.3) \\ &= E\left(-R_t \mid -R_t > -x_{t,\alpha} = VaR_\alpha\right). \end{aligned}$$

The result below gives an expression for the ES in terms of the probability density function (p.d.f.) of the return R_t .

Result 3.1.1

Let $f_{R_t}(r)$ be the p.d.f. of the return distribution, then

$$ES_\alpha = -\alpha^{-1} \int_{-\infty}^{x_{t,\alpha}} r f_{R_t}(r) dr.$$

Proof:

The conditional distribution of R_t given that $R_t < x_{t,\alpha}$ is given by

$$\begin{aligned} f_{R_t|x_{t,\alpha}}(r | R_t < x_{t,\alpha}) &= \frac{f_{R_t}(r)}{P(R_t < x_{t,\alpha})} \\ &= \frac{f_{R_t}(r)}{\alpha}. \end{aligned}$$

Hence,

$$\begin{aligned} ES_\alpha &= -E(R_t | R_t < x_{t,\alpha}) \\ &= -\int_{-\infty}^{x_{t,\alpha}} r f_{R_t|x_{t,\alpha}}(r | R_t < x_{t,\alpha}) dr \\ &= -\alpha^{-1} \int_{-\infty}^{x_{t,\alpha}} r f_{R_t}(r) dr. \quad \blacksquare \end{aligned}$$

This section has provided some basic definitions and results concerning Value at Risk and Expected Shortfall. In the next section some of the disadvantages of VaR are discussed and the need for an alternative risk measure such as Expected Shortfall is explained.

3.2. Comparison of VaR and ES

Although commonly used, VaR has some disadvantages as a risk measure. Firstly, VaR is not subadditive (McNeil, 1999). A risk measure $R(\cdot)$ is said to be sub-additive if

$$R(c_1 + c_2 + \dots + c_m) \leq R(c_1) + R(c_2) + \dots + R(c_m)$$

where $c_i, i = 1, \dots, m$ are the m constituents of the portfolio (Ruppert, 2004). The VaR of the total portfolio may exceed the sum of the constituent VaRs. Secondly, VaR does not measure the extent of exceptional losses (Alexander, 2008(b)).

These disadvantages have led to the consideration of alternative risk measures, of which ES (also referred to as expected tail loss, or conditional VaR) has become the preferred metric to be used for regulatory and economic capital allocation. The ES risk measure is sub-additive and better captures the extent of exceptional losses. The VaR ranking and the ES rankings for return series may be different, indicating the usefulness of ES as an alternative risk measure (Gilli *et al.*, 2006).

Traditional VaR measures use historical data as an indicator for what may happen in the future. This is not necessarily a good indicator, especially in extreme market events (Deloitte Market Risk). E.g. in 2008, the VaR measures understated the risk. After the fall of Lehman Brothers in 2008, the historical volatility estimates were extremely high, leading to an overstatement of risk in 2009. An article by Deloitte entitled *Market Risk* (<http://www.deloitte.com/assets/Dcom-SouthAfrica/Local%20Assets/Documents/7.%20Basel%20flyer%20-%20Market%20Risk.pdf>), suggests that ES and Extreme Value theory (EVT) can be useful augmentations to VaR. EVT is discussed in detail in chapter 3.

The most straight-forward application of VaR and ES assumes independent, identically distributed returns. This is the subject of the next section.

3.3. Models for independent, identically distributed returns

Two models are considered. The first assumes that the returns are independent identically distributed (i.i.d.) according to a Normal distribution. The second assumes that the returns follow an i.i.d. generalised Student t distribution. In both of these models, the mean and variance are constants, which may not be realistic in practice.

3.3.1. Normal model

If $R_t \sim i.i.d. N(\mu, \sigma^2)$ it is trivial to derive the VaR_α , i.e.:

$$\begin{aligned}
 P(R_t < x_{t,\alpha}) &= \alpha \\
 \Rightarrow P\left(Z_t = \frac{R_t - \mu}{\sigma} < \frac{x_{t,\alpha} - \mu}{\sigma}\right) &= \alpha; \quad Z_t \sim N(0,1) \\
 \Rightarrow \frac{x_{t,\alpha} - \mu}{\sigma} &= \Phi^{-1}(\alpha) = -\Phi^{-1}(1 - \alpha) \\
 \Rightarrow x_{t,\alpha} &= \mu - \sigma \cdot \Phi^{-1}(1 - \alpha) \\
 \Rightarrow VaR_\alpha = -x_\alpha &= \sigma \cdot \Phi^{-1}(1 - \alpha) - \mu, \quad (3.4)
 \end{aligned}$$

with $\Phi^{-1}(\cdot)$ the inverse of the standard normal cumulative distribution function (c.d.f.).

From the definition of ES in definition 3.1.3 and from result 3.1.1, it can also be shown that

$$ES_{\alpha} = \alpha^{-1} \phi(\Phi^{-1}(\alpha)) \sigma - \mu, \quad (3.5)$$

with $\varphi(\cdot)$ the standard normal p.d.f.. This result can be derived as follows:

Firstly, the ES of $Z_t \sim N(0,1)$ is derived with $Z \equiv Z_t$ in the integral.

$$\begin{aligned} ES_{\alpha}(Z_t) &= -\alpha^{-1} \int_{-\infty}^{\Phi^{-1}(\alpha)} z \phi(z) dz \\ &= -\alpha^{-1} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\Phi^{-1}(\alpha)} z \exp\left(-\frac{1}{2} z^2\right) dz \\ &= \alpha^{-1} \frac{1}{\sqrt{2\pi}} \left[\exp\left(-\frac{1}{2} z^2\right) \right]_{-\infty}^{\Phi^{-1}(\alpha)} \\ &= \alpha^{-1} \frac{1}{\sqrt{2\pi}} \left[\exp\left(-\frac{1}{2} (\Phi^{-1}(\alpha))^2\right) - 0 \right] \\ &= \alpha^{-1} \phi(\Phi^{-1}(\alpha)) \end{aligned}$$

In result 3.1.1 make the transformation $R_t = \mu + \sigma Z_t$ so that $R_t \sim N(\mu, \sigma^2)$, $dr = \sigma dz$,

$$r = x_{\alpha} \Rightarrow z = \frac{x_{\alpha} - \mu}{\sigma} = \Phi^{-1}(\alpha) \text{ and } r = -\infty \Rightarrow z = -\infty.$$

Hence

$$\begin{aligned} ES_{\alpha}(R_t) &= -\alpha^{-1} \int_{-\infty}^{x_{\alpha}} r f_{R_t}(r) dr \\ &= -\alpha^{-1} \int_{-\infty}^{\Phi^{-1}(\alpha)} (z\sigma + \mu) \underbrace{f_{R_t}(\mu + \sigma z)}_{\phi(z)} \sigma dz \\ &= -\alpha^{-1} \int_{-\infty}^{\Phi^{-1}(\alpha)} z \phi(z) \sigma dz - \alpha^{-1} \mu \int_{-\infty}^{\Phi^{-1}(\alpha)} \phi(z) dz \\ &= -\alpha^{-1} [-\phi(\Phi^{-1}(\alpha)) \sigma] - \alpha^{-1} \mu \alpha \\ &= \alpha^{-1} \phi(\Phi^{-1}(\alpha)) \sigma - \mu \end{aligned}$$

3.3.2. Student t model

Suppose R_t follows a generalised Student t distribution with v degrees of freedom, expected value of μ and variance σ^2 then along the same lines as in section 3.3.1, it can be derived that

$$VaR_{\alpha,v} = \sqrt{v^{-1}(v-2)}t_v^{-1}(1-\alpha)\sigma - \mu \quad (3.6)$$

with $t_v^{-1}(1-\alpha)$ the $(1-\alpha)^{\text{th}}$ quantile of the standard Student t distribution (Alexander, 2008(b)).

In this case the Expected Shortfall is given by:

$$ES_{\alpha,v} = \alpha^{-1}(v-1)^{-1}(v-2+x_\alpha(v)^2)f_v(x_\alpha(v))\sigma - \mu \quad (3.7)$$

where $x_\alpha(v)$ is the α^{th} quantile of the generalised t distribution $f_v(\cdot)$ (Alexander, 2008(b)).

3.3.3. The problem with the i.i.d. assumption

A major problem in the above models is that it is assumed that the mean (μ) and variance (σ^2) are constant over time. Therefore the forecasts from these models equal their estimates. However, in practice the i.i.d. assumption is unrealistic, since the mean and variance of returns are time dependent. This is especially so in daily returns because of the volatility clustering effect (Alexander, 2008(a)).

One way to deal with this problem is to model the conditional variance equation and the conditional mean equation separately. This is achieved by GARCH conditional models that treat the mean and variance as time-varying functions. Model estimates are then obtained and used in the calculation of VaR.

3.4. The GARCH conditional model

AR-GARCH conditional models are used to account for the time varying nature of the mean and the variance of returns. A distinction is made between a Normal conditional error distribution and a standardised Student t conditional error distribution. In the following sections, a detailed model description is provided, maximum likelihood estimation is discussed, and lastly forecasts and long term volatility are considered.

3.4.1. Model description

A GARCH model consists of two equations: a conditional mean equation which specifies the behaviour of the returns and a conditional variance equation which describes the dynamic behaviour of the conditional variance.

Definition 3.4.1: GARCH conditional mean equation

The conditional mean equation for a GARCH model is given by:

$$R_t = \mu_t + E_t$$

where

$$E_t = \sigma_t Z_t$$

and

$$Z_t \sim N(0,1) \text{ or } Z_t \sim \text{standardised Student t distribution.} \quad \blacksquare$$

From this it follows that

$$E(R_t) = \mu_t \text{ and } \text{Var}(R_t) = \sigma_t^2 .$$

For example, if R_t follows a AR(1) process, i.e. $\mu_t = \varphi_0 + \varphi_1 R_{t-1}$, the conditional mean equation is given by

$$R_t = \varphi_0 + \varphi_1 R_{t-1} + E_t \quad (3.8)$$

Other examples of the process of R_t are the AR(p), MA(q) and ARMA(p,q) processes.

Definition 3.4.2: Symmetric GARCH(m,n) model

The symmetric GARCH(m,n) model for describing the conditional variance is given by:

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_n \varepsilon_{t-n}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_m \sigma_{t-m}^2. \quad \blacksquare$$

The GARCH(1,1) model given by

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (3.9)$$

is the most frequently used GARCH model. In this model:

- The parameters satisfy the constraints $\omega > 0$; $\alpha, \beta \geq 0$ and $(\alpha + \beta) < 1$.
- The quantity ε_{t-1} is the observed value of the random variable

$$E_{t-1} = R_{t-1} - \mu_{t-1}$$

which is estimated by

$$\varepsilon_{t-1} \approx r_{t-1} - \bar{r}$$

with r_{t-1} the observed return at time $t - 1$ and \bar{r} the mean of the n historical observed returns up to time $t - 1$, i.e.

$$\bar{r} = \frac{1}{n} \sum_{i=1}^n r_{t-i}.$$

It is important to note that σ_t^2 is modelled using all the information (all the past returns) up to and including time $t - 1$. Hence

$$(R_t - \mu_t) | I_{t-1} = E_t | I_{t-1} \sim N(0, \sigma_t^2) \text{ or}$$

~ generalised Student t distribution with expected value 0

and variance σ_t^2

or

$$R_t | I_{t-1} \sim N(\mu_t, \sigma_t^2) \text{ or}$$

~ generalised Student t distribution with expected value μ_t and variance σ_t^2 .

The random variable E_t is also called the mean corrected return, or the random shock. The distribution of $E_t | I_{t-1}$ is conditional on all the information available up to time $t-1$. The dynamic behaviour of the variance is therefore accounted for by E_t . The dependency of the E_t random shocks on the past is reflected in the definition of σ_t^2 which states that today's variance is a function of yesterday's variance plus a random shock (realisation of the mean corrected return). The process is not identically distributed or independent, because the conditional variances at different time points are related.

The parameters in the conditional variance equation (2.9) can be interpreted as follows. The constant parameter ω is related to the frequency of returns (typically small for daily returns). The error parameter α indicates the sensitivity of the reaction of the conditional variance to market shocks. The lag parameter β indicates persistence in conditional volatility, irrespective of market events.

The Student t distribution is a useful alternative to the normal distribution, since the Student t distribution allows for excess kurtosis that is often encountered in the conditional distribution of daily returns. The maximum likelihood estimation depends on whether a normal distribution or a Student t distribution is used. This is the subject of the next section.

3.4.2. Maximum likelihood estimation

The parameters in the GARCH(1,1) model are estimated by maximum likelihood. A distinction is made between the case where $E_t | I_{t-1} \sim N(0, \sigma_t^2)$ and the case where $E_t | I_{t-1}$ is Student t distributed with degrees of freedom ν , expected value 0 and variance σ_t^2 .

3.4.2.1. Normal model

If $E_t|I_{t-1} \sim N(0, \sigma_t^2)$, the parameters in (2.8), denoted by $\theta = (\omega, \alpha, \beta)^T$, can be estimated by maximising the normal likelihood function:

$$L^N(\theta) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi}\sigma_t} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right) \quad (3.10)$$

which is equivalent to maximising:

$$\ell^N(\theta) = \sum_{t=1}^T \left[-\ln(\sigma_t) - \frac{\varepsilon_t^2}{\sigma_t^2}\right]. \quad (3.11)$$

Note the dependency of the likelihood function on $\theta = (\omega, \alpha, \beta)^T$ since σ_t^2 in (3.9) depends on these parameters.

The following algorithm is used for the parameter estimation:

Algorithm 3.4.1: Maximum likelihood estimation for the Normal GARCH(1,1) model

Suppose a set of observed portfolio prices p_t , $t=1,2, \dots, T$ are available.

1. Calculate $r_t = \ln p_t - \ln p_{t-1}$, $t = 2, 3, \dots, T$.

$$\text{Take } \bar{r}_1 = \frac{1}{T-1} \sum_{t=2}^T r_t \triangleq \bar{r}.$$

2. Calculate $\varepsilon_t^2 = (r_t - \bar{r})^2$, $t = 2, 3, \dots, T$.

$$\text{Take } \varepsilon_1^2 = \frac{1}{T-1} \sum_{t=2}^T \varepsilon_t^2.$$

3. Decide on starting values for ω , α and β . Call these ω_0 , α_0 and β_0 .

Note that ω is related to the frequency of the data and will typically be small for daily data (Alexander, 2008(a)).

4. Calculate the GARCH conditional variances:

$$\text{For } t = 2: \quad \sigma_2^2 = \omega_0 + \alpha_0 \varepsilon_1^2 + \beta_0 \varepsilon_1^2;$$

$$\text{for } t = 3, 4, \dots, T: \quad \sigma_t^2 = \omega_0 + \alpha_0 \varepsilon_{t-1}^2 + \beta_0 \sigma_{t-1}^2.$$

5. Calculate the term for maximisation:

$$\ell^N(\theta) = \sum_{t=2}^T \left[-\ln(\sigma_t) - \frac{\varepsilon_t^2}{\sigma_t^2}\right]. \text{ Note the dependency of } \ell^N(\theta) \text{ on the starting values}$$

ω_0 , α_0 and β_0 .

6. $\ell^N(\theta)$ is maximised by varying the input values of ω_0 , α_0 and β_0 (subject to the constraints $\omega > 0$; $\alpha, \beta \geq 0$ and $\alpha + \beta < 1$).

Steps 3 to 5 are repeated, until a maximum value is obtained. Various search methods can be implemented for which any optimisation software can be used. Most computer packages use the Levenberg-Marquardt algorithm for optimisation (Alexander, 2008(a)). The values $\hat{\omega}$, $\hat{\alpha}$ and $\hat{\beta}$ are such that $\ell^N(\hat{\theta}) = \max(\ell^N(\theta))$ are the maximum likelihood estimates.

3.4.2.2. Student t model

If $E_t|I_{t-1}$ follows a Student t distribution, then the degrees of freedom must also be estimated by maximum likelihood. The conditional variance equation does not change, but the likelihood function will be different.

Result 3.4.1: Log-likelihood for the Student t GARCH(1,1) model

Let $E_t|I_{t-1}$ follow a Student t distribution with degrees of freedom ν , expected value 0 and variance σ_t^2 and let $\theta = (\omega, \alpha, \beta)^T$. Then the log-likelihood function is (Alexander, 2008(a)):

$$\begin{aligned} \ell^t(\theta) &= \log L^t(\theta) \\ &= -\sum_{t=1}^T \left\{ \log(\sigma_t) + \left(\frac{\nu+1}{2}\right) \log\left(1 + \frac{\left(\frac{\varepsilon_t}{\sigma_t}\right)^2}{(\nu-2)}\right) \right\} + T \log \left\{ \left((\nu-2)\pi\right)^{-\frac{1}{2}} \Gamma\left(\frac{\nu}{2}\right)^{-1} \Gamma\left(\frac{\nu+1}{2}\right) \right\} \end{aligned}$$

■

Proof (Alexander, 2008(a)):

The return at time t is E_t with $E(E_t) = 0$ and $Var(E_t) = \sigma_t^2$ where

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2.$$

The standardised return at time t is $\frac{E_t}{\sigma_t}$ with $E\left(\frac{E_t}{\sigma_t}\right) = 0$ and $Var\left(\frac{E_t}{\sigma_t}\right) = 1$.

Assume that $\frac{E_t}{\sigma_t}$ has a standardised- t distribution

$$f_\nu(t) = \left((\nu-2)\pi\right)^{-\frac{1}{2}} \Gamma\left(\frac{\nu}{2}\right)^{-1} \Gamma\left(\frac{\nu+1}{2}\right) \left(1 + \frac{t^2}{(\nu-2)}\right)^{-\left(\frac{\nu+1}{2}\right)}$$

with expected value of 0 and a variance of 1.

The unknown GARCH parameters in σ_t must be estimated by maximum likelihood. The ε_t values are observed for $t = 1, 2, \dots, T$. To write down the likelihood function, the distribution of E_t must be known and the parameters of σ_t must be contained in the likelihood function to enable their estimation.

The p.d.f. of E_t is obtained from $f_v(t)$ via the transformation $t = \frac{E_t}{\sigma_t}$:

$$\begin{aligned} f_v(E_t) &= f_v\left(\frac{E_t}{\sigma_t}\right) \left| \frac{\partial t}{\partial E_t} \right|; \frac{\partial t}{\partial E_t} = \frac{1}{\sigma_t} \\ &= ((v-2)\pi)^{-\frac{1}{2}} \Gamma\left(\frac{v}{2}\right)^{-1} \Gamma\left(\frac{v+1}{2}\right) \left(1 + \frac{\left(\frac{E_t}{\sigma_t}\right)^2}{(v-2)}\right)^{-\left(\frac{v+1}{2}\right)} \frac{1}{\sigma_t}. \end{aligned}$$

Letting $\theta = (\omega, \alpha, \beta)^T$, the likelihood function is

$$L^t(\theta) = \prod_{t=1}^T \left\{ ((v-2)\pi)^{-\frac{1}{2}} \Gamma\left(\frac{v}{2}\right)^{-1} \Gamma\left(\frac{v+1}{2}\right) \left(1 + \frac{\left(\frac{\varepsilon_t}{\sigma_t}\right)^2}{(v-2)}\right)^{-\left(\frac{v+1}{2}\right)} \frac{1}{\sigma_t} \right\}$$

and the log likelihood is

$$\begin{aligned} \ell^t(\theta) &= \log L^t(\theta) \\ &= -\sum_{t=1}^T \left\{ \log(\sigma_t) + \left(\frac{v+1}{2}\right) \log\left(1 + \frac{\left(\frac{\varepsilon_t}{\sigma_t}\right)^2}{(v-2)}\right) \right\} + T \log \left\{ ((v-2)\pi)^{-\frac{1}{2}} \Gamma\left(\frac{v}{2}\right)^{-1} \Gamma\left(\frac{v+1}{2}\right) \right\}. \blacksquare \end{aligned}$$

Having derived the log-likelihood function of the Student t GARCH model the parameters can be estimated using the following algorithm.

Algorithm 3.4.2: Maximum likelihood estimation for the Student t GARCH(1,1) model

Let $E_t | I_{t-1}$ follow a Student t distribution with degrees of freedom v , expected value 0 and variance σ_t^2 . Suppose a set of observed portfolio prices p_t , $t=1, 2, \dots, T$ are available.

1. Calculate $r_t = \ln p_t - \ln p_{t-1}$, $t = 2, 3, \dots, T$.

$$\text{Take } r_1 = \frac{1}{T-1} \sum_{t=2}^T r_t \triangleq \bar{r}.$$

2. Calculate $\varepsilon_t^2 = (r_t - \bar{r})^2$, $t = 2, 3, \dots, T$.

$$\text{Take } \varepsilon_1^2 = \frac{1}{T-1} \sum_{t=2}^T \varepsilon_t^2.$$

3. Decide on starting values for ω , α , β and ν . Call these ω_0 , α_0 , β_0 and ν_0 .

Note that ω is related to the frequency of the data and will typically be small for daily data (Alexander, 2008(a)).

4. Calculate the GARCH conditional variances:

$$\text{For } t = 2: \quad \sigma_2^2 = \omega_0 + \alpha_0 \varepsilon_1^2 + \beta_0 \varepsilon_1^2;$$

$$\text{for } t = 3, 4, \dots, T: \quad \sigma_t^2 = \omega_0 + \alpha_0 \varepsilon_{t-1}^2 + \beta_0 \sigma_{t-1}^2.$$

5. Calculate the term for maximisation:

$$\ell^t(\boldsymbol{\theta}) = -\sum_{t=1}^T \left\{ \log(\sigma_t) + \left(\frac{\nu+1}{2}\right) \log\left(1 + \frac{\left(\frac{\varepsilon_t}{\sigma_t}\right)^2}{(\nu-2)}\right) \right\} + T \log\left\{ ((\nu-2)\pi)^{-\frac{1}{2}} \Gamma\left(\frac{\nu}{2}\right)^{-1} \Gamma\left(\frac{\nu+1}{2}\right) \right\}.$$

Note the dependency of $\ell^t(\boldsymbol{\theta})$ on the starting values ω_0 , α_0 , β_0 and ν_0 .

6. $\ell^t(\boldsymbol{\theta})$ is maximised by varying the input values of ω_0 , α_0 , β_0 and ν_0 (subject to the constraints $\omega > 0$; $\alpha, \beta \geq 0$, $\alpha + \beta < 1$ and $\nu > 0$).

Steps 3 to 5 are repeated, until a maximum value is obtained. The values $\hat{\omega}$, $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\nu}$ are such that $\ell^t(\hat{\boldsymbol{\theta}}) = \max(\ell^t(\boldsymbol{\theta}))$ are the maximum likelihood estimates.

These time-varying estimates of the mean, variance and degrees of freedom are substituted into the equations for VaR and ES in (3.6) and (3.7) to obtain different, dynamic estimates each day. The maximum likelihood estimates derived in this section are applied in the next section concerning forecasts and long term volatility of a GARCH model.

3.4.3. Forecasts and long term volatility

The GARCH volatility in equation (3.9), i.e.

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

converges to a long term unconditional volatility which is derived in result 3.4.2 below. Next, the one day ahead forecasts from the model are considered. These are different from the estimated volatility, because the returns are not identically distributed. Using the one day ahead forecasts, a GARCH volatility forecast term structure can be obtained. Finally, it is shown that this term structure converges to the long term volatility.

3.4.3.1. Long term volatility

The GARCH volatility converges to the following long term unconditional volatility (Alexander, 2008(a)):

Result 3.4.2: Long term volatility of GARCH(1,1) model

$$\bar{\sigma}^2 = \frac{\omega}{1-(\alpha+\beta)}.$$

Proof:

In equation (3.9)

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

under a long term assumption, substitute

$$\sigma_t^2 \approx \sigma_{t-1}^2 \approx \bar{\sigma}^2$$

and

$$\varepsilon_{t-1}^2 \approx \bar{\sigma}^2,$$

since

$$E(\varepsilon_{t-1}^2) = \sigma_{t-1}^2 = \bar{\sigma}^2.$$

This yields:

$$\bar{\sigma}^2 = \omega + \alpha \bar{\sigma}^2 + \beta \bar{\sigma}^2 \Rightarrow \bar{\sigma}^2 = \frac{\omega}{1-(\alpha+\beta)}.$$

3.4.3.2. One day ahead forecasts

The forecasts from the GARCH(1,1) model are not equal to the estimates from the GARCH model. This is what distinguishes the GARCH model from the models of section 3.3 that assume independent, identically distributed returns.

Result 3.4.3: One day ahead forecasts for the GARCH(1,1) model

Let $\hat{\omega}$, $\hat{\alpha}$ and $\hat{\beta}$ be the maximum likelihood estimates. The one day ahead forecasts for the GARCH(1,1) model from day $T + s$ to day $T + s + 1$ is given by (Alexander, 2008(a)) as:

$$\hat{\sigma}_{T+s+1}^2 = \hat{\omega} + (\hat{\alpha} + \hat{\beta})\hat{\sigma}_{T+s}^2 . \quad \blacksquare$$

Proof:

Suppose day T is the last day in the sample. All information up to and including day T is available and denoted by I_T .

Consider the estimated model

$$\hat{\sigma}_t^2 = \hat{\omega} + \hat{\alpha}\varepsilon_{t-1}^2 + \hat{\beta}\hat{\sigma}_{t-1}^2, \quad t=1, 2, \dots, T$$

where

$$\varepsilon_{t-1}^2 \approx (r_{t-1} - \bar{r})^2$$

and

$$\hat{\sigma}_{t-1}^2 = \hat{\omega} + \hat{\alpha}\varepsilon_{t-2}^2 + \hat{\beta}\hat{\sigma}_{t-2}^2 .$$

Therefore ε_{t-1} and $\hat{\sigma}_{t-1}$ are calculated from the information up to and including time $t - 1$ and hence assumed to be non-stochastic for $t=1, 2, \dots, T$.

The one day ahead forecast for time $(T+1)$, seen at time T is

$$\hat{\sigma}_{T+1}^2 = \hat{\omega} + \hat{\alpha}\varepsilon_T^2 + \hat{\beta}\hat{\sigma}_T^2$$

with

$$\varepsilon_T^2 = (r_T - \bar{r})^2$$

the last residual in the GARCH model and

$$\hat{\sigma}_T^2 = \hat{\omega} + \hat{\alpha}\varepsilon_{T-1}^2 + \hat{\beta}\hat{\sigma}_{T-1}^2 .$$

Both these quantities are non-stochastic and fully defined by the information set I_T .

The one day ahead forecast for time $(T+2)$, seen at time $(T+1)$, is

$$\hat{\sigma}_{T+2}^2 = \hat{\omega} + \hat{\alpha}E_{T+1}^2 + \hat{\beta}\hat{\sigma}_{T+1}^2$$

where E_{T+1}^2 is an unknown random variable at time T .

Since

$$E(E_{T+1}^2) = \sigma_{T+1}^2 \Rightarrow E_{T+1}^2 \approx \hat{\sigma}_{T+1}^2,$$

the following approximation holds:

$$\hat{\sigma}_{T+2}^2 \approx \hat{\omega} + \hat{\alpha}\hat{\sigma}_{T+1}^2 + \hat{\beta}\hat{\sigma}_{T+1}^2 = \hat{\omega} + (\hat{\alpha} + \hat{\beta})\hat{\sigma}_{T+1}^2 .$$

A recursive argument can be developed along the same lines as the above. The forecasted forward daily variance from day $(T + s)$ to day $(T + s + 1)$ is then easily seen to be given by

$$\hat{\sigma}_{T+s+1}^2 = \hat{\omega} + (\hat{\alpha} + \hat{\beta})\hat{\sigma}_{T+s}^2$$

or in terms of σ_{T+1}^2 follows

$$\sigma_{T+s+1}^2 = \omega + (\alpha + \beta)^s \sigma_{T+1}^2. \quad \blacksquare$$

The one day ahead forecasts are therefore time varying. The next section discusses this term structure.

3.4.3.3. Term structure of volatility forecasts

The volatility forecasts from the GARCH model lead to volatility forecasts that are not constant, but have a term structure.

The daily forecasts can be used to obtain a h -day term structure of volatility forecasts (Alexander, 2008(a)). A ten day forecast is often needed in VaR calculations. To obtain this ten day forecasted volatility, the average over the one day forecasts made at time T , time $T + 1$, etc. up to time $T + 9$ is calculated. This amounts to setting $S=0, 1, \dots, 9$ in the daily forecasting formula $\hat{\sigma}_{T+s+1}^2 = \hat{\omega} + (\hat{\alpha} + \hat{\beta})\hat{\sigma}_{T+s}^2$ (result 3.4.3). The average of these values is the 10 day forecast.

The GARCH term structure forecasts are mean reverting. This is discussed in the next section.

3.4.3.4. Convergence of forecasts to long term volatility

The GARCH volatility forecasts will in the long term converge to a stabilising value. As the next result shows, this stabilising value is exactly equal to the long term unconditional volatility of result 3.4.2.

Result 3.4.4: Convergence of forecasts to long term volatility GARCH(1,1) model

The GARCH(1,1) forecasts of the variance converges to the long-term average variance as the time of the forecast increases (Alexander, 2008(a)):

$$\hat{\sigma}_{T+s+1}^2 = \hat{\omega} + (\hat{\alpha} + \hat{\beta})\hat{\sigma}_{T+s}^2 \rightarrow \bar{\sigma}^2 = \frac{\hat{\omega}}{1-(\hat{\alpha}+\hat{\beta})} \text{ if } s \rightarrow \infty. \quad \blacksquare$$

Proof:

$$\begin{aligned} \hat{\sigma}_{T+s+1}^2 &= \hat{\omega} + (\hat{\alpha} + \hat{\beta})\hat{\sigma}_{T+s}^2 \\ &= \hat{\omega} + (\hat{\alpha} + \hat{\beta})(\hat{\omega} + (\hat{\alpha} + \hat{\beta})\hat{\sigma}_{T+s-1}^2) \\ &= \hat{\omega} + (\hat{\alpha} + \hat{\beta})\hat{\omega} + (\hat{\alpha} + \hat{\beta})^2\hat{\sigma}_{T+s-1}^2 \\ &= \hat{\omega} + (\hat{\alpha} + \hat{\beta})\hat{\omega} + (\hat{\alpha} + \hat{\beta})^2(\hat{\omega} + (\hat{\alpha} + \hat{\beta})\hat{\sigma}_{T+s-2}^2) \\ &= \dots \\ &= \hat{\omega} \left(1 + (\hat{\alpha} + \hat{\beta}) + (\hat{\alpha} + \hat{\beta})^2 + \dots + (\hat{\alpha} + \hat{\beta})^{s-1} \right) + (\hat{\alpha} + \hat{\beta})^s \hat{\sigma}_{T+1}^2 \\ &= \hat{\omega} \frac{1-(\hat{\alpha}+\hat{\beta})^s}{1-(\hat{\alpha}+\hat{\beta})} + (\hat{\alpha} + \hat{\beta})^s \hat{\sigma}_{T+1}^2 \rightarrow \frac{\hat{\omega}}{1-(\hat{\alpha}+\hat{\beta})} = \bar{\sigma}^2 \text{ if } s \rightarrow \infty, \\ &\text{since } (\hat{\alpha} + \hat{\beta}) < 1. \quad \blacksquare \end{aligned}$$

3.5. Asymmetric GARCH conditional models

Asymmetric GARCH conditional models are a popular variation on the symmetric models in section 3.4. The reason for this is explained next. The rest of the section discusses some popular asymmetric GARCH models. The asymmetric GARCH (A-GARCH) and the GJR-GARCH models have an extra parameter to enable an asymmetric response. The exponential GARCH (E-GARCH) includes an asymmetric response function.

3.5.1. The rationale for asymmetric models

The GARCH models of the previous section are symmetric in the sense that positive and negative returns have the same effect on the GARCH volatility. However, in equity markets, the volatility usually increases more after a large negative return, than it does after a large positive return (Alexander, 2008(a)). This is due to the leverage effect: a decrease in the

stock price leads to an increase in the debt-equity ratio of a firm, causing the firm to be more highly leveraged and the stock price to be more volatile. There is therefore a negative correlation between equity returns and volatility. Consequently, GARCH models that respond asymmetrically to positive and negative returns are considered. The results and discussions in the rest of this section follow from Alexander (2008(a)).

3.5.2. A-GARCH and GJR-GARCH

The A-GARCH and GJR-GARCH models differ from the symmetric GARCH model in that an additional parameter, lambda (λ), is included. This allows the conditional variance to be affected differently depending on the sign of the return.

3.5.2.1. Model definition, long term variance and forecasts

In the table below, the A-GARCH (asymmetric GARCH) and the GJR-GARCH are compared in terms of model definition, long term variance and forecasts.

Table 3.5.1: Comparison of A-GARCH and GJR-GARCH

A-GARCH	GJR-GARCH
<i>Model definition</i>	
$\sigma_t^2 = \omega + \alpha(\varepsilon_{t-1} - \lambda)^2 + \beta\sigma_{t-1}^2.$	$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \lambda I\{\varepsilon_{t-1} < 0\}\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2,$ Where $I\{\varepsilon_{t-1} < 0\} = 1$, if $\varepsilon_{t-1} < 0$ $= 0$, elsewhere.
<i>Long-term variance</i>	
$\bar{\sigma}^2 = \frac{\omega + \lambda^2\alpha}{1 - (\alpha + \beta)}$	$\bar{\sigma}^2 = \frac{\omega}{1 - (\alpha + \beta + \frac{1}{2}\lambda)}$
<i>One step ahead forecast</i>	
$\hat{\sigma}_{T+1}^2 = \hat{\omega} + \hat{\alpha}(\varepsilon_T - \hat{\lambda})^2 + \hat{\beta}\sigma_T^2.$	$\hat{\sigma}_{T+1}^2 = \hat{\omega} + \hat{\alpha}\varepsilon_T^2 + \hat{\lambda}I\{\varepsilon_T < 0\}\varepsilon_T^2 + \hat{\beta}\sigma_T^2.$
<i>S step ahead forecast</i>	
$\hat{\sigma}_{T+S+1}^2 = (\hat{\omega} + \hat{\lambda}^2\hat{\alpha}) + (\hat{\alpha} + \hat{\beta})\hat{\sigma}_T^2.$	$\hat{\sigma}_{T+S+1}^2 = \hat{\omega} + (\hat{\alpha} + \hat{\beta} + \frac{1}{2}\hat{\lambda})\hat{\sigma}_T^2.$

The additional parameter λ captures the leverage effect. In the A-GARCH model, if $\lambda > 0$, the effect on the conditional variance will be larger for a negative return. If $\lambda < 0$ then a positive return will have a greater effect. In the GJR-GARCH, the term $\lambda I\{E_{t-1} < 0\}E_{t-1}^2$ only changes the variance response from symmetric to asymmetric if a negative return is observed. The A-GARCH and the GJR-GARCH yield very similar results in practice (Alexander, 2008(a)).

3.5.2.2. Estimation

The parameters in the above models can be estimated by maximum likelihood. The procedure is similar to that described for the symmetric GARCH models in section 3.4.2, except that a starting value λ_0 for the additional parameter λ need also to be selected. The expression for σ_t^2 differs according to the model and the likelihood for maximisation also includes λ via σ_t^2 (although the likelihood has the same form as previously, depending on whether normal or Student t innovations are assumed).

3.5.3. E-GARCH

In the E-GARCH (Exponential GARCH) model a conditional equation for the *log* variance is considered. This ensures a positive variance without the need of parameter constraints. An asymmetric response function is included in the model definition that responds differently to positive and negative returns.

3.5.3.1. Model definition

The model definition (Alexander, 2008(a)) is:

Definition 3.5.1: The E-GARCH model

Let the return at time t be

$$R_t = c + \sigma_t Z_t,$$

with

$$Z_{t-1} \sim i. i. d. N(0,1).$$

The E-GARCH model for σ_t is defined as

$$\log(\sigma_t^2) = \omega + g(Z_{t-1}) + \beta \log(\sigma_{t-1}^2), \quad (3.12)$$

with the asymmetric response function given by

$$g(Z_t) = \theta Z_t + \gamma(|Z_t| - \sqrt{2/\pi}). \quad \blacksquare$$

3.5.3.2. Properties

The following list provides some important properties of the E-GARCH model.

1. If $E_t = \sigma_t Z_t$, the E-GARCH model corresponds to a Normal GARCH(1,1) model with mean equation

$$R_t = c + E_t.$$

2. The function $g(\cdot)$ is the asymmetric response function. The term

$$\gamma(|Z_t| - \sqrt{2/\pi})$$

signifies the difference between an observed $|Z_t|$ and its expected value

$$E(|Z_t|) = \sqrt{2/\pi}, \quad Z_{t-1} \sim N(0,1).$$

3. The response function reacts differently to positive and negative values of Z_t . The

function is linear with a y-axis intercept of $-\sqrt{2/\pi}$, but the slope will be

$$(\theta + \gamma) \text{ if } Z_t > 0$$

and

$$(\theta - \gamma) \text{ if } Z_t < 0.$$

4. The choice of θ allows for a variety of asymmetric responses. If $\theta = \gamma$, then only positive returns will cause a response. If $\theta = -\gamma$, then the conditional log variance will only be influenced by negative returns.

3.5.3.3. Long term variance

The long-term variance forecast from the E-GARCH model will eventually settle down to a steady state value (Alexander, 2008(a)).

Result 3.5.1: Long term variance for the E-GARCH model

$$\bar{\sigma}^2 = \exp\left(\frac{\omega}{1-\beta}\right). \quad \blacksquare$$

Proof:

In equation (3.12), i.e.

$$\log(\sigma_t^2) = \omega + g(Z_{t-1}) + \beta \log(\sigma_t^2),$$

let

$$\bar{\sigma}^2 \approx \sigma_t^2 \approx \sigma_{t-1}^2.$$

Also note that

$$\begin{aligned} E[g(Z_t)] &= \theta E[Z_t] + \gamma \left(E[|Z_t|] - \sqrt{2/\pi} \right) \\ &= \theta(0) + \gamma \left(\sqrt{2/\pi} - \sqrt{2/\pi} \right) \\ &= 0, \end{aligned}$$

since $Z_{t-1} \sim N(0,1)$.

Hence $g(Z_{t-1})$ is approximated by:

$$g(Z_{t-1}) \approx 0.$$

Therefore

$$\begin{aligned} \log(\bar{\sigma}^2) &\approx \omega + 0 + \beta \log(\bar{\sigma}^2) \\ \Rightarrow (1-\beta) \log(\bar{\sigma}^2) &\approx \omega \\ \Rightarrow \log(\bar{\sigma}^2) &\approx \frac{\omega}{1-\beta} \\ \Rightarrow \bar{\sigma}^2 &\approx \exp\left(\frac{\omega}{1-\beta}\right). \end{aligned} \quad \blacksquare$$

3.5.3.4. Estimation

The parameters of the E-GARCH model can be estimated by maximum likelihood. The steps are provided in Algorithm 3.5.1 if $Z_{t-1} \sim i.i.d.N(0,1)$. Algorithm 3.5.2 considers the maximisation if the distribution is not normal, but Student t.

Algorithm 3.5.1: Maximum likelihood estimation for the Normal E-GARCH(1,1) model

Suppose a set of observed portfolio prices p_t , $t=1,2, \dots,T$ are available.

1. Calculate $r_t = \ln p_t - \ln p_{t-1}$, $t = 2,3,\dots,T$.
Take $\bar{r}_1 = \frac{1}{T-1} \sum_{t=2}^T r_t \triangleq \bar{r}$.
2. Calculate $\varepsilon_t^2 = (r_t - \bar{r})^2$, $t = 2,3,\dots,T$.
Take $\varepsilon_1^2 = \frac{1}{T-1} \sum_{t=2}^T \varepsilon_t^2$.
3. Decide on starting values for ω , α , β and γ . Call these ω_0 , α_0 , β_0 and γ_0 .
Choose some value for $\log(\sigma_1^2)$ (Alexander (2008(a)) suggests -10 in an example).
4. Set $Z_1 = \frac{\varepsilon_1}{\sigma_1}$, with $\sigma_1 = \sqrt{\exp(\log(\sigma_1^2))}$.
5. Calculate $g(Z_1) = \theta_0 Z_1 + \gamma_0 (|Z_1| - \sqrt{2/\pi})$.
6. Repeat the next steps ($t = 2,3,\dots,T$):
 - a. $\log(\sigma_t^2) = \omega + g(Z_{t-1}) + \beta \log(\sigma_{t-1}^2)$
 - b. $\sigma_t = \sqrt{\exp(\log(\sigma_t^2))}$
 - c. $Z_t = \frac{\varepsilon_t}{\sigma_t}$
 - d. $g(Z_t) = \theta Z_t + \gamma (|Z_t| - \sqrt{2/\pi})$
7. Calculate the term for maximisation $\ell^N(\omega, \theta, \gamma, \beta) = \sum_{t=1}^T [-\ln(\sigma_t) - \frac{\varepsilon_t^2}{\sigma_t^2}]$.

Steps 6 and 7 are repeated for different starting values. Those set of starting values that maximise $\ell^N(\omega, \theta, \gamma, \beta)$ are the maximum likelihood estimates.

Definition 3.5.1. can be adjusted so that $Z_{t-1} \sim$ standardised Student t with ν degrees of freedom, expected value of zero and variance of one. This implies that $E_t | I_{t-1} \sim$ Student t with degrees of freedom ν , expected value zero and variance σ_t^2 . Steps 3 and 7 of algorithm 3.5.1 can then be adjusted to yield algorithm 3.5.2 for maximum likelihood estimation of the parameters.

Algorithm 3.5.2: Maximum likelihood estimation for the Student t E-GARCH(1,1) model

Adjust the following steps of algorithm 3.5.1:

- Step 3:
Decide on a starting value for the degrees of freedom, say ν_0 , in addition to the other starting values.
- Step 7:
The term for maximisation now equals

$$\ell^t(\omega, \theta, \gamma, \beta, \nu) = -\sum_{t=1}^T \left\{ \log(\sigma_t) + \left(\frac{\nu+1}{2}\right) \log\left(1 + \frac{\left(\frac{\varepsilon_t}{\sigma_t}\right)^2}{(\nu-2)}\right) \right\} + T \log\left\{ \left((\nu-2)\pi\right)^{-\frac{1}{2}} \Gamma\left(\frac{\nu}{2}\right)^{-1} \Gamma\left(\frac{\nu+1}{2}\right) \right\}.$$

3.5.3.5. One day forecasts

The variance forecasts from the E-GARCH model are non-constant. The result and its proof follow from Alexander (2008(a)).

Result 3.5.2: One day ahead forecasts from the E-GARCH model

Suppose day T is the last day in the sample, then the forecast for day $T+1$, made on day T is

$$\hat{\sigma}_{T+1}^2 = \exp(\hat{\omega}) \exp(g(Z_T)) (\hat{\sigma}_T^2)^{\hat{\beta}}.$$

In general, for $S=1,2,\dots$, the forecast for day $T+S+1$, made on day $T+S$, is

$$\hat{\sigma}_{T+S+1}^2 = \exp\left(\hat{\omega} - \hat{\gamma} \sqrt{2/\pi}\right) \cdot \hat{c} \cdot \hat{\sigma}_{T+S}^{2\hat{\beta}},$$

with

$$\hat{c} = \exp\left(\frac{1}{2}(\hat{\gamma} + \hat{\theta})^2\right) \Phi(\hat{\gamma} + \hat{\theta}) + \exp\left(\frac{1}{2}(\hat{\gamma} - \hat{\theta})^2\right) \Phi(\hat{\gamma} - \hat{\theta}). \quad \blacksquare$$

Proof:

Consider equation (3.12):

$$\log(\sigma_t^2) = \omega + g(Z_{t-1}) + \beta \log(\sigma_{t-1}^2).$$

Taking logs on both sides, gives

$$\sigma_t^2 = \exp(\omega) \exp(g(Z_{t-1})) (\sigma_{t-1}^2)^\beta.$$

Take $t = T + S + 1$ so that

$$\hat{\sigma}_{T+S+1}^2 = \exp(\hat{\omega}) E[\exp(g(Z_{T+S})) | I_{T+S}] \hat{\sigma}_{T+S}^{2\hat{\beta}}. \quad (3.13)$$

Simplify $E[\exp(g(Z_{T+S})) | I_{T+S}]$ in (3.13):

$$\begin{aligned} E[\exp(g(Z_{T+S})) | I_{T+S}] &= E[\exp(\hat{\theta}Z_{T+S} + \hat{\gamma}(|Z_{T+S}| - \sqrt{2/\pi}))] \\ &= \int_{-\infty}^{\infty} \exp\left(\hat{\theta}z + \hat{\gamma}\left(|z| - \sqrt{2/\pi}\right)\right) \varphi(z) dz \\ &= \exp\left(-\hat{\gamma}\sqrt{2/\pi}\right) \left\{ \int_{-\infty}^0 \exp\left(z(\hat{\theta} - \hat{\gamma})\right) \varphi(z) dz + \right. \\ &\quad \left. \int_0^{\infty} \exp\left(z(\hat{\theta} + \hat{\gamma})\right) \varphi(z) dz \right\}. \end{aligned} \quad (3.14)$$

To solve (3.14), consider:

$$\begin{aligned} \int_0^{\infty} \exp(ax) \varphi(x) dx &= \frac{1}{\sqrt{2\pi}} \int_0^{\infty} \exp(ax) \exp\left(-\frac{1}{2}x^2\right) dx \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(\frac{1}{2}a^2\right) \int_0^{\infty} \exp\left(-\frac{1}{2}(x-a)^2\right) dx \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(\frac{1}{2}a^2\right) \int_{-a}^{\infty} \exp\left(-\frac{1}{2}y^2\right) dy \quad (y = x - a) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(\frac{1}{2}a^2\right) \sqrt{2\pi}(1 - \Phi(-a)) \\ &= \exp\left(\frac{1}{2}a^2\right) \Phi(a). \end{aligned}$$

Similarly:

$$\int_{-\infty}^0 \exp(ax) \varphi(x) dx = \exp\left(\frac{1}{2}a^2\right) \Phi(-a)$$

Therefore (3.14) follows as

$$\begin{aligned} E[\exp(g(Z_{T+S})) | I_{T+S}] &= \exp\left(-\hat{\gamma}\sqrt{2/\pi}\right) \left\{ \exp\left(\frac{1}{2}(\hat{\gamma} + \hat{\theta})^2\right) \Phi(\hat{\gamma} + \hat{\theta}) + \right. \\ &\quad \left. \exp\left(\frac{1}{2}(\hat{\gamma} - \hat{\theta})^2\right) \Phi(\hat{\gamma} - \hat{\theta}) \right\} \\ &= \exp\left(-\hat{\gamma}\sqrt{2/\pi}\right) \cdot \hat{c} \quad \blacksquare \end{aligned}$$

3.6 Summary

In this chapter the VaR and ES measures as they are traditionally applied in risk management were considered. Basic definitions and results for VaR and ES were given. Models assuming independent, identically distributed returns were discussed, with a distinction made between a normal distribution and a Student t distribution for the returns. Symmetric GARCH conditional models that allow for time variation in the conditional mean and variance equations received detailed attention, with the focus also on maximum likelihood estimation, forecasts and long term volatility. The fact that the GARCH volatilities have a term structure that is mean-reverting was highlighted. Finally, some asymmetric GARCH models were discussed that allow volatility to be affected differently by positive and negative returns.

The drawback of the approaches above is that they do not account for heavy tails. This issue can be addressed by augmenting the traditional theory with Extreme Value Theory. Extreme Value Theory is the subject of the next chapter.

4. Traditional EVT

4.1 Introduction

Extreme risks are situated in the tails of probability distributions. Extreme Value Theory (EVT) is concerned with the modelling of this tail behaviour. Pareto-type distributions are fat-tailed and therefore a suitable class of alternative distributions to the traditional assumptions of symmetric Normal or Student t distributed returns. Since financial data are mostly fat-tailed, these traditional models may underestimate both the size and frequency of extreme risks. Traditional models work well in the centre of the distribution of the data, but are less efficient in the tails of distributions where the data is sparse.

One of the main goals of EVT is the estimation of extreme quantiles. This makes EVT models a natural choice for a VaR or ES model. EVT allows tail estimation and extrapolation beyond observed data points to estimate very high quantiles (Fernandez, 2003).

Throughout this paper, let $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ denote the order statistics of data X_1, X_2, \dots, X_n from a distribution with distribution function $F(x) = P(X \leq x)$.

Definition 4.1.1: Quantile and tail quantile functions

The quantile function of a random variable $X \sim F$ is defined as

$$Q(p) := \inf\{x : F(x) \geq p\} \quad (4.1)$$

and the tail quantile function as

$$U(x) = Q\left(1 - \frac{1}{x}\right), \quad (4.2)$$

where $Q(\cdot) = F^{-1}(\cdot)$. ■

Remark:

Typically p will be very small, leading to the estimation of very high quantiles in regions of sparse data, or no data at all.

Extreme Value theory has two main approaches. The older approach is the so-called Block Maxima approach. The sample is divided into blocks of equal length and the maximum in each block is determined. If the required conditions are satisfied, then the limiting distribution for the block maxima can be obtained. The second approach is the Peaks over Threshold (POT) approach that has come to be preferred in recent years. This involves choosing some threshold return and then only considering returns above this threshold value. More detail on the Block Maxima and the POT approaches are provided in sections 4.2 and section 4.3 respectively.

In section 4.3 the focus is extensively on the Fréchet class of distributions, since that class contains the Pareto-type distributions that are of particular interest for this research. The Hill estimator as well as threshold selection methods will be considered, followed by an in-depth discussion of quantile estimation. In section 4.4 it is explained how the POT approach is used to determine VaR and ES. In section 4.5 some limitations of EVT are provided.

4.2 Method of Block Maxima

The method of Block Maxima consists of dividing the sample into blocks of equal length and then determining the maximum in each block. If appropriately standardised, the maximum follows a limiting GEV (Generalised Extreme Value) distribution.

Suppose X_1, X_2, \dots, X_n are independent and identically distributed according to a distribution function F . Let the maximum of the sample be denoted by

$$X_{nn} = \max(X_1, X_2, \dots, X_n).$$

Then for appropriate choices of the constants a_n and b_n (i.e. if such constants exist), Fisher and Tippett (1928), Gnedenko (1943) and De Haan (1970) proved that the only possible limiting distribution for the maximum is the Generalised Extreme Value distribution (GEV, denoted by G_γ below).

Definition 4.2.1: GEV as limiting distribution for block maxima

Under appropriate conditions on the underlying distribution, and if constants $\{a_n > 0\}$ and

$\{b_n\}$ exist such that $\frac{X_{nm} - b_n}{a_n}$ converge in distribution, then

$$\frac{X_{nm} - b_n}{a_n} \xrightarrow{D} G_\gamma(x) = \exp(-(1 + \gamma x)^{-1/\gamma}) \text{ for } (1 + \gamma x) > 0 \text{ and } \gamma \in \mathbb{R}. \quad \blacksquare$$

In general, the GEV distribution includes a location parameter, μ , and a scale parameter, σ , in addition to the shape parameter, γ . The general definition of a GEV distribution is:

Definition 4.2.2: GEV distribution

$$G_{\sigma, \gamma, \mu}(x) = \begin{cases} \exp\left(-\left(1 + \gamma \frac{x - \mu}{\sigma}\right)^{-1/\gamma}\right), & 1 + \gamma \frac{x - \mu}{\sigma} > 0, \gamma \neq 0 \\ \exp\left(-\exp\left(-\frac{x - \mu}{\sigma}\right)\right), & x \in \mathfrak{R}, \gamma = 0. \end{cases} \quad \text{with } \sigma > 0 \text{ and } \mu \in \mathfrak{R}$$

■

The parameters of the GEV can be estimated by maximum likelihood methods using the following result:

Result 4.2.1: Log-likelihood for observations from a GEV distribution

Suppose x_1, x_2, \dots, x_m is a sample from a GEV distribution, then the log likelihood is given by

$$\log L(\sigma, \gamma, \mu) = \begin{cases} -m \log \sigma - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^m \log\left(1 + \gamma \frac{x_i - \mu}{\sigma}\right) - \sum_{i=1}^m \left(1 + \gamma \frac{x_i - \mu}{\sigma}\right)^{-1/\gamma}, & \gamma \neq 0 \\ -m \log \sigma - \sum_{i=1}^m \exp\left(-\frac{x_i - \mu}{\sigma}\right) - \sum_{i=1}^m \frac{x_i - \mu}{\sigma}, & \gamma = 0. \end{cases}$$

■

The maximum likelihood estimates have a joint asymptotic normal distribution if $\gamma > -0.5$ of the form

$$\sqrt{m}((\hat{\sigma}, \hat{\gamma}, \hat{\mu}) - (\sigma, \gamma, \mu)) \xrightarrow{D} N(0, V_1)$$

where V_1 is the inverse of the Fisher information matrix. See Beirlant *et al.* (2004) for details.

There are three classes of GEV distributions, depending on the value of the shape parameter. The first class is the Fréchet class (also known as the Fréchet-Pareto class) corresponding to $\gamma > 0$. Distributions falling in this class are also termed Pareto-type distributions and they have heavy tails, i.e. exhibit HTE (heavier tail than exponential) behaviour. Examples include the strict Pareto, the Generalised Pareto, the Fréchet and Burr distributions. The second class is obtained if $\gamma = 0$ and is called the Gumbel class with examples including the Weibull, Exponential, Gamma, Logistic and Log-normal distributions. The final class is the Weibull class that corresponds to the case $\gamma < 0$. In this case, the distributions have light tails, i.e. exhibit LTE (lighter tail than exponential) behaviour.

It is clear that the Fréchet class ($\gamma > 0$) is of interest when determining VaR and ES measures, since financial series are fat-tailed. In the discussion in section 4.4 the focus is on Pareto-type EVT distributions.

One very useful application of the Block Maxima approach, is the calculation of return levels. The return level is defined as a quantile of the GEV distribution and it is the maximum return that we expect will be exceeded once in every k periods of length n . The definition is as follows (Gilli *et al.*, 2006):

Definition 4.2.3: Estimated return level (Gilli *et al.*, 2006)

$$\begin{aligned} \hat{R}_n^k &= G_{\hat{\sigma}, \hat{\gamma}, \hat{\mu}}^{-1}\left(1 - \frac{1}{k}\right) \\ &= \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\gamma}} \left(1 - \left(-\log\left(1 - \frac{1}{k}\right)\right)\right) \\ \hat{\mu} - \hat{\sigma} \log\left(-\log\left(1 - \frac{1}{k}\right)\right), \hat{\gamma} = 0. \end{cases} \end{aligned}$$

■

The return level is a more conservative measure than VaR (Gilli *et al.*, 2006). The GEV parameters in definition 4.2.2 are estimated using maximum likelihood and the log-likelihood function of result 4.2.1.

Although the Block Maxima approach is used for the determination of return levels as described above, it is not often used in other practical applications. The POT method (refer section 4.3) is preferred.

According to Fernandez (2003) there are two reasons for this preference. The first is that the block maxima method may lead to a loss of information, since only the maxima are extracted. Other information may be obtained by considering the rest of the sample as well. The second reason is that it is easier to compute VaR and ES estimates based on the POT approach (refer section 4.5).

Another disadvantage of the Block Maxima method is uncertainty regarding the choice of the block length. Although similar in nature to the problem of selecting a threshold in POT models, for POT models there exist certain guidelines for threshold selection (albeit not decisive). POT models are the subject of the next section.

4.3 Peaks over Threshold (POT) approach

It is clear that using only the block maxima for model-building is not making efficient use of all the information provided by the data. As an alternative to the Block Maxima method, the POT (Peaks over Threshold) approach has been developed and will be outlined in this section.

The POT approach considers the absolute exceedances above a sufficiently high threshold. Suppose the original data is X_1, X_2, \dots, X_n and denote the threshold value by u . Exceedances of the form

$$\{Y_j = X_i - u \text{ given } X_i > u\}_{j=1}^{N_u} \quad (4.3)$$

are considered where N_u is the number of exceedances.

The distribution of these exceedances, for a sufficiently high threshold, can be approximated by a Generalised Pareto distribution (Beirlant *et al.*, 2004). The motivation for the GPD will be discussed next, followed by the formal definition of a GPD and the issue of parameter estimation.

4.3.1 Theoretical motivation for using the GPD

The theoretical justification can be seen from the condition C_γ^* as stated in Beirlant *et al.* (2004). This is a necessary and sufficient condition that must be satisfied by a distribution function F so that it is in the domain of attraction of the GEV distribution (i.e. so that $F \in D(G_\gamma)$). This means that if F satisfies C_γ^* , then the limiting distribution of the maximum will be GEV (refer definition 4.2.3). The condition is stated formally as follows in the Pickands, Balkema and de Haan Theorem (Beirlant *et al.*, 2004):

Result 4.3.1: The Pickands, Balkema and de Haan Theorem

The distribution F belongs to $D(G_\gamma)$ if and only if for some auxiliary function b and $1 + \gamma y > 0$,

$$\frac{1 - F(t + b(t)y)}{1 - F(t)} \rightarrow (1 + \gamma y)^{-\frac{1}{\gamma}} \text{ as } t \rightarrow \infty,$$

then

$$\frac{b(t + yb(t))}{b(t)} \rightarrow u^\gamma = 1 + \gamma y.$$

■

To see how this theorem justifies the GPD, consider:

$$\begin{aligned} \frac{1 - F(t + b(t)y)}{1 - F(t)} &= \frac{P(X > t + b(t)y)}{P(X > t)} \\ &= \frac{P(X - t > b(t)y)}{P(X - t > 0)} \quad (\text{now take } Y = X - t) \\ &= P(Y > yb(t) | Y > 0) \\ &= 1 - F_t(yb(t)). \end{aligned}$$

From C_γ^* it follows that

$$1 - F_t(yb(t)) \approx (1 + \gamma y)^{-\frac{1}{\gamma}},$$

or equivalently that

$$1 - F_t(y) \approx (1 + \gamma \frac{y}{b(t)})^{-\frac{1}{\gamma}}.$$

Setting $b(t) = \sigma$ leads to the GPD as an approximation to the distribution of the exceedances. The definition of the GPD is provided in the next section.

4.3.2 Generalised Pareto Distribution (GPD)

The distribution function and the density functions are given in the next two definitions. In the rest of this section various estimators for the parameters, as well as confidence intervals, are discussed.

Definition 4.3.1: Generalised Pareto Distribution function

$$H_{\gamma,\sigma}(y) = \begin{cases} 1 - \left(1 + \frac{\gamma y}{\sigma}\right)^{-\frac{1}{\gamma}}, & y \in (0, \infty), \gamma > 0, \\ 1 - \exp\left(-\frac{y}{\sigma}\right), & y \in (0, \infty), \gamma = 0, \\ 1 - \left(1 + \frac{\gamma y}{\sigma}\right)^{-\frac{1}{\gamma}}, & y \in \left(0, -\frac{\sigma}{\gamma}\right), \gamma < 0. \end{cases}$$

■

Here γ is the extreme value index (EVI) and we expect it to be positive, since the distribution of the returns is positively skewed. The case where $\gamma > 0$ will receive further attention.

Definition 4.3.2: Generalised Pareto density function ($\gamma > 0$)

$$h_{\gamma,\sigma}(y) = \frac{1}{\sigma} \left(1 + \frac{\gamma y}{\sigma}\right)^{-\frac{1}{\gamma} - 1}.$$

■

4.3.2.1 Parameter estimation

The parameters of the GPD can be estimated in various ways. Maximum likelihood is perhaps the most popular, but method of moments estimators are also used.

For the maximum likelihood method, re-parameterise by setting $\tau = \frac{\gamma}{\sigma}$ and then maximise the following log-likelihood:

Result 4.3.2: Log-likelihood for observations from a GPD if $\gamma > 0$

Suppose y_1, y_2, \dots, y_{N_u} follow a GPD distribution, then

$$L(\tau, \gamma) = \log l(\tau, \gamma) = -N_u \log(\gamma) + N_u \log(\tau) - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^{N_u} \log(1 + \tau y_i). \quad \blacksquare$$

The method of moments estimators involve equating the population moments to their sample estimates. Suppose Y_1, Y_2, \dots, Y_{N_u} follow a GPD distribution, then equate

$$E(Y) = \frac{\sigma}{1-\gamma} \text{ to } \bar{Y} = \frac{1}{N_u} \sum_{i=1}^{N_u} Y_i$$

and

$$Var(Y) = \frac{\sigma^2}{(1-\gamma)^2(1-2\gamma)} \text{ to } S_Y^2 = \frac{1}{N_u-1} \sum_{i=1}^{N_u} (Y_i - \bar{Y})^2.$$

This yields the following estimates:

Result 4.3.3: Method of moments estimators for the parameters of the GPD

$$\hat{\gamma}_{MOM} = \frac{1}{2} \left(1 - \frac{\bar{Y}^2}{S_Y^2} \right) \text{ and } \hat{\sigma}_{MOM} = \frac{\bar{Y}}{2} \left(1 + \frac{\bar{Y}^2}{S_Y^2} \right). \quad \blacksquare$$

The method of moments estimation can be adjusted slightly by considering *probability weighted* moments. The so-called probability weighted moment of a random variable Y with distribution function F is defined by Greenwood *et al.* (1979) as:

Definition 4.3.3: Probability Weighted Moment

The (p, r, s) th probability weighted moment is defined as

$$M_{p,r,s} = E\{Y^p [F(Y)]^r [1 - F(y)]^s\} \text{ for real } p, r \text{ and } s.$$

(See e.g. Hosking *et al.*, 1985.) \blacksquare

The procedure described in Beirlant *et al.* (2004) is followed for the estimation of the Generalised Pareto parameters. This involves setting $p = 1$, $r = 0$ and $s = 0, 1$ and equating the population moment $M_{p,r,s}$ to the corresponding sample estimate, $\hat{M}_{p,r,s}$, given by

$$M_{p,r,s} = \frac{1}{N_u} \sum_{j=1}^{N_u} Y_{j,N_u} \left(\frac{j}{N_u+1} \right)^r \left(1 - \frac{j}{N_u+1} \right)^s.$$

Since $M_{1,0,0} = \frac{\sigma}{1-\gamma}$ and $M_{1,0,1} = \frac{\sigma}{2(2-\gamma)}$, the following estimators are obtained:

Result 4.3.4: Method of probability weighted moments estimator for the parameters of the GPD

$$\hat{\gamma}_{PVM} = 2 - \frac{\bar{M}_{1,0,0}}{\bar{M}_{1,0,0} - 2\bar{M}_{1,0,1}} \text{ and } \hat{\sigma}_{PVM} = \frac{2\bar{M}_{1,0,0}\bar{M}_{1,0,1}}{\bar{M}_{1,0,0} - 2\bar{M}_{1,0,1}}.$$

■

The maximum likelihood, method of moments and method of probability weighted moments estimators all have an asymptotic normal distribution. This has the general form of

$$\sqrt{N_u}((\hat{\gamma}, \hat{\sigma}) - (\gamma, \sigma)) \xrightarrow{D} N(0, V)$$

where $\hat{\gamma}$ and $\hat{\sigma}$ are the specific estimators and V is a covariance matrix that will depend on the type of estimator considered. For example, for the maximum likelihood estimates (if $\gamma > -0.5$),

$$V = V_1 = (1 + \gamma) \begin{bmatrix} 1 + \gamma & -\sigma \\ -\sigma & 2\sigma^2 \end{bmatrix}. \tag{4.4}$$

The expressions for V_2 and V_3 are given by Beirlant *et al.* (2004).

Approximate $100(1-\alpha)\%$ confidence intervals can be constructed. They will have the form below:

Result 4.3.5: Approximate $100(1-\alpha)\%$ confidence intervals for parameters of GPD

$$\hat{\gamma} \pm \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \sqrt{\frac{\hat{v}_{1,1}}{N_u}} \text{ and } \hat{\sigma} \pm \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \sqrt{\frac{\hat{v}_{2,2}}{N_u}} \tag{4.5}$$

where $\hat{\gamma}$ and $\hat{\sigma}$ are the specific estimators and $\hat{v}_{1,1}$ and $\hat{v}_{2,2}$ are the first and second diagonal elements of either V_1 , V_2 or V_3 , as appropriate.

■

Good alternatives to the above approximate normal confidence intervals are profile log-likelihood confidence intervals, based on the profile likelihood ratio statistic (Beirlant *et al.*, 2004). Let $L(\gamma, \sigma)$ be the log-likelihood for the GPD (result 4.3.2 with $\tau = \frac{Y}{\sigma}$), then the profile likelihood function for γ is found by maximising the log-likelihood, conditional on γ . Denote this by

$$L^*(\gamma) = \max_{\sigma|\gamma} L(\gamma, \sigma),$$

then the $100(1-\alpha)\%$ profile likelihood confidence interval for γ is given by the following result:

Result 4.3.6: Approximate $100(1-\alpha)\%$ profile log-likelihood confidence intervals for the parameters of the GPD

$$\{\gamma : \log L^*(\gamma) \geq \log L^*(\hat{\gamma}) - \frac{\chi_1^2(1-\alpha)}{2}\}. \quad \blacksquare$$

In this section the POT approach to EVT was discussed, as well as the Generalised Pareto distribution that plays a central role in this approach. Various estimators for the parameters were touched upon. Attention will now focus on detail surrounding distributions in the Fréchet class.

4.4 Fréchet class of distributions

The Fréchet class corresponds to the first class of GEV distributions where $\gamma > 0$. Distributions in this class are also called Pareto-type distributions. As mentioned earlier, these distributions have heavy tails. More specifically, if $x \rightarrow \infty$ then the survival function $1-F(x)$ tends to zero almost at polynomial speed, apart from the slowly varying component.

To define a Pareto-type distribution, the tail quantile function and functions of regular variation must first be defined. These basic definitions will be provided in the next section.

4.4.1 Basic Definitions

Functions of regular variation and slow variation play important roles in EVT models.

Definition 4.4.1: Functions of regular variation

A function $A(\cdot)$ is said to be regular varying of order γ if

$$\frac{A(xt)}{A(x)} \rightarrow t^\gamma, \quad x \rightarrow \infty, \quad \gamma \in \mathfrak{R}. \quad \blacksquare$$

In the special case where $\gamma = 0$ and $\frac{\ell(xt)}{\ell(x)} \rightarrow 1, \quad x \rightarrow \infty, \quad \ell(\cdot)$ is said to be slowly varying.

With the knowledge of tail quantile functions and regularly and slowly varying functions, the Pareto-type distribution can now be defined. The definition can either be in terms of the tail quantile function (refer definition 4.1.1, equation (4.2)), or in terms of the underlying distribution.

Definition 4.4.2: Pareto type distributions

Pareto type distributions are distributions for which the tail quantile function is regularly varying of order γ , i.e.:

$$U(x) = x^\gamma \ell_U(x)$$

with $\ell_U(x)$ slowly varying. ■

Remarks:

Definition 4.4.2 implies that

$$\frac{U(xt)}{U(x)} = \frac{(xt)^\gamma \ell_U(xt)}{(x)^\gamma \ell_U(x)} \rightarrow t^\gamma, \quad x \rightarrow \infty, \quad \text{for all } t > 0.$$

Alternatively, Pareto type distributions are distributions for which the survival function is regular varying of order $-\frac{1}{\gamma}$:

$$1 - F(x) = x^{-\frac{1}{\gamma}} \ell_F(x),$$

with $\ell_F(x)$ slowly varying, implying that

$$\frac{1-F(xt)}{1-F(x)} = \frac{(xt)^{-\frac{1}{\gamma}} \ell_F(xt)}{(x)^{-\frac{1}{\gamma}} \ell_F(x)} \rightarrow t^{-\frac{1}{\gamma}}, \quad x \rightarrow \infty. \quad (4.6)$$

The De Bruyn conjugate provides a one-to-one relationship between ℓ_U and ℓ_F (Beirlant *et al.*, 2004).

The implication of (4.6) is that if the original distribution function F of the data is Pareto-type, then the relative exceedances follow an approximate Pareto distribution (if the slowly varying part is ignored and the threshold is high enough). To see why this is so, denote the

relative exceedances by $Y = \frac{X}{t} | X > t$. Then

$$\begin{aligned} P(Y > x) &= P\left(\frac{X}{t} > x | X > t\right) \\ &= P(X > tx | X > t) \\ &= \frac{P(X > tx)}{P(X > t)} \\ &= \frac{1-F(xt)}{1-F(x)} \rightarrow t^{-\frac{1}{\gamma}}, \quad x \rightarrow \infty. \end{aligned}$$

So far, some preliminary definitions of Pareto-type distributions and the theory surrounding it were given. For these distributions the parameter $\gamma > 0$. This parameter is called the extreme value index and plays a very important role in EVT.

4.4.2. The Extreme Value Index

The extreme value index (EVI) is the shape parameter γ in the GEV distribution (definition 4.2.2) and the GPD (definition 4.3.1). It characterises the tail behaviour of the extreme value distribution relative to that of an exponential distribution, for which $\gamma = 0$. For fat-tailed financial series, the focus is on the Fréchet class of distributions that exhibits HTE (heavier tail than exponential) behaviour and has $\gamma > 0$.

The most popular estimator of the EVI is the Hill estimator (Hill, 1975).

Definition 4.4.3: The Hill estimator

$$\hat{\gamma}_{Hill} = H_{k,n} = \frac{1}{k} \sum_{j=1}^k \log X_{n-j+1,n} - \log X_{n-k,n}. \quad (4.7) \blacksquare$$

In definition 4.4.3, k is the number of exceedances above the threshold, taken as the order statistic $X_{n-k,n}$. It represents the number of values in the sample that is considered as extreme enough for estimation of the EVI.

The following distributional properties of the Hill estimator are useful.

Result 4.4.1: Distributional properties of the Hill estimator (Beirlant *et al.*, 2004)

- i. Asymptotic variance of the Hill estimator:

$$\text{AVar}(\hat{\gamma}_{Hill}) = \frac{\gamma^2}{k}. \quad (4.8)$$

- ii. Asymptotic normal distribution:

$$\sqrt{k} \left(\frac{\hat{\gamma}_{Hill}}{\gamma} - 1 \right) \xrightarrow{d} N(0,1).$$

- iii. (1- α) Confidence intervals for the Hill estimator:

$$\left(\hat{\gamma}_{Hill} - \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \frac{\hat{\gamma}}{\sqrt{k}}; \hat{\gamma}_{Hill} + \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \frac{\hat{\gamma}}{\sqrt{k}} \right),$$

or

$$\left(\hat{\gamma}_{Hill} \left(1 + \frac{\Phi^{-1} \left(1 - \frac{\alpha}{2} \right)}{\sqrt{k}} \right)^{-1}; \hat{\gamma}_{Hill} \left(1 - \frac{\Phi^{-1} \left(1 - \frac{\alpha}{2} \right)}{\sqrt{k}} \right)^{-1} \right).$$

The second confidence interval is expected to lead to better coverage (Beirlant *et al.*, 2004). \blacksquare

The Hill estimator is only appropriate if $\gamma > 0$. A generalisation of the Hill estimator that is often used if to estimate $\gamma \in \mathbb{R}$, is the moment estimator (Dekkers *et al.*, 1989; Beirlant *et al.*, 2004). The definition is given below.

Definition 4.4.4: The Moment estimator

$$M_{k,n} = H_{k,n} + 1 - \frac{1}{2} \left(1 - \frac{H_{k,n}^2}{H_{k,n}^{(2)}} \right)^{-1}$$

where

$$H_{k,n}^{(2)} = \frac{1}{k} \sum_{j=1}^k (\log X_{n-j+1,n} - \log X_{n-k,n})^2. \quad \blacksquare$$

This section has discussed the well-known Hill estimator for the extreme value index, which should be positive for the Fréchet class. Attention is now turned again toward the POT approach. As was argued previously, if the underlying distribution function is in the Fréchet class, then the POT approach implies that the exceedances over a sufficiently high threshold approximately follow a GPD. The choice of the threshold is very important and the selection thereof, as well as exploratory data analysis (to assess the validity of the Fréchet class assumption), are discussed next.

4.4.3. Threshold selection and exploratory data analysis

The threshold in the POT models are generally set equal to the $(n-k)^{\text{th}}$ order statistic, i.e. set $u = X_{n-k,n}$. This allows consideration of exactly k exceedances. The problem, however, is the choice of k . The variance-bias trade-off is relevant here. The larger the value of k , the more observations will be used to construct parameter and quantile estimates. The estimates will consequently have lower variance. However, for large values of k , the analysis looks deeper into the order statistics, where the limiting results of Extreme Value theory may not hold and where the slowly varying part of the Pareto-type distribution is not negligible. The information contained in these data points may therefore not be relevant. As a result, the estimators will tend to have large bias.

Various methods have been developed to aid this choice, but unfortunately there is no prescribed “correct method”. Much of this process is trial and error. McNeil and Frey (2000) suggest taking 5-10% of the sample as extreme. Naturally, the quantile estimates will be affected by the choice of the threshold. Therefore it is often sensible to construct plots of the change in the quantile estimates as a function of the threshold.

A few graphical methods commonly used in threshold selection will be discussed briefly (QQ plots, Hill plots and mean excess plots). The threshold selection methods may be simultaneously used to explore the nature of the data. For example, the fat-tailed assumption can be validated and it can be ascertained whether an extreme value theory model is appropriate for the data at hand.

4.4.3.1. Quantile-Quantile (QQ) plots

In the tail, the Pareto-type distribution is approximated by a Pareto distribution with parameter equal to the EVI (if the slowly varying component is ignored). The justification is as follows.

For a Pareto-type distribution, the tail quantile function is of the form $U(x) = x^\gamma \ell_U(x)$ (definition 4.4.2). This means that

$$\begin{aligned}\log U(x) &= \gamma \log x + \log \ell(x) \\ \Rightarrow \frac{\log U(x)}{\log x} &= \gamma + \frac{\log \ell(x)}{\log x} \\ \Rightarrow \frac{\log U(x)}{\log x} &\approx \gamma, \quad x \rightarrow \infty.\end{aligned}$$

since

$$\frac{\log \ell(x)}{\log x} \rightarrow 0, \quad x \rightarrow \infty \text{ (Bierlant *et al.*, 2004).}$$

Therefore $\log U(x) \approx \gamma \log x$ for large x .

Furthermore, for an Exponential(λ) distribution, the tail quantile function is of the form

$$U_E(x) = -\frac{1}{\lambda} \log x = -\gamma \log x \text{ (set } \gamma = \frac{1}{\lambda}\text{).}$$

Since the tail quantile functions are similar, it is expected that the logs of the upper order statistics form the Pareto-type distribution, will behave like order statistics from an Exponential distribution.

The Exponential QQ plot of the observed quantiles against the theoretical (exponential) quantiles involves plotting the points $(\log X_{n-j+1,n}; -\log \frac{j}{n+1})$, $(j = 1, \dots, k)$ and fixing at the threshold point $(\log X_{n-k,n}; -\log \frac{k+1}{n+1})$ (Beirlant *et al.*, 2004).

Exponential QQ plots have three uses. The first use is to assess the goodness of fit. The validity of a Pareto-type distribution to the excess returns can be checked. If appropriate, the QQ plot will be roughly linear. The second use is that the EVI ($\gamma = \frac{1}{\lambda}$) can be estimated as the slope of the QQ plot. Last, but not least, the QQ plot is used for threshold selection. The value of k should be taken as that value from which the QQ plot is roughly linear.

4.4.3.2. Hill plots

Hill plots are another way to determine the threshold for GPDs. The Hill plot involves plotting the Hill estimators (definition 4.4.3) against k . I.e.,

$$\{(k, H_{k,n}): 1 \leq k \leq n-1\}.$$

Usually, k is chosen in a region where the plot seems constant. Unfortunately Hill plots are rarely constant and judgement is very subjective. For this reason, Hill plots are often referred to as Hill “horror” plots (Beirlant *et al.*, 2004).

4.4.3.3. Mean excess plots

The mean excess plot can also be used to determine the level of the threshold and to check that it is reasonable to assume that the underlying distribution falls in the Fréchet domain. See also for example Beirlant *et al.* (2004) for details.

The mean excess function is defined as

$$e(u) = E(X - u | X > u) = E(X | X > u) - u. \quad (4.9)$$

For the GPD, the mean excess function is (Bensalah, 2000):

$$e(u) = \frac{\sigma + \gamma u}{1 - \gamma}, \quad \sigma + \gamma u > 0. \quad (4.10)$$

This function is linear and increasing. The Exponential distribution has the constant mean excess function $e(u) = \frac{1}{\lambda}$.

The mean excess function can be estimated as (Beirlant *et al.*, 2004):

$$\hat{e}_n(t) = \frac{\frac{1}{n} \sum_{i=1}^n X_i \text{Ind}(X_i > t)}{\frac{1}{n} \sum_{i=1}^n \text{Ind}(X_i > t)} - t. \quad (4.11)$$

For the mean excess plots, set $t = X_{n-k,n}$ in (4.11), yielding:

$$e_{k,n} = \hat{e}_n(t = X_{n-k,n}) = \frac{1}{k} \sum_{j=1}^k X_{n-j+1,n} - X_{n-k,n}.$$

Two types of mean excess plots are constructed (over a range of k values deemed as possible threshold choices):

1. Plot $e_{k,n}$ against k and consider small k values (look from the right to the left of the plot).
2. Plot $e_{k,n}$ against $\log(x_{n-k,n})$ and consider large $\log(x_{n-k,n})$ values (look from the left to the right of the plot).

If the behaviour of the above two plots are increasing in the direction as indicated, then evidence for the heavier tail than exponential (HTE) property of the distribution is provided. The point from which the increasing and linear behaviour starts, can be a rough indication of an appropriate threshold choice.

This section has discussed the non-trivial problem of threshold selection. Various graphical methods, such as QQ plots, Hill plots and mean excess plots can be used to aid the selection process and to validate the Fréchet class assumption. In the next section the important aspect of quantile estimation will be discussed.

4.5 Quantile estimation

The estimation of extreme quantiles is of utmost important in EVT. It is necessary for both VaR and ES calculations, since these have directly to do with extreme quantiles. The

quantile function was defined in (4.1). Both first order and second order estimation will be discussed.

4.5.1 First order estimation

In this section first order estimation of quantiles is considered. Firstly, the Weissmann (1978) estimator based on the Hill estimate of the EVI is discussed. Next, the so-called probability view estimator will be derived. This inverts the Generalised Pareto distribution to obtain quantile estimators.

4.5.1.1 The Weissman estimator

First order estimators ignore the slowly varying part of Pareto-type distributions. The Weissman (1978) estimator is a first order quantile estimator that uses the Hill estimator of the EVI and is based on linear regression of the Pareto quantile plot. The result is given below, followed by an outline of the proof.

Result 4.5.1: Weissman (1978) estimator for extreme quantiles

$$\hat{q}_{k,p}^+ = \hat{Q}(1-p) = X_{n-k,n} \left(\frac{k+1}{(n+1)p} \right)^{H_{k,n}}.$$

Here p is some small probability. ■

Proof:

The proof is based on the linear regression of a Pareto quantile plot. Assume a strict Pareto distribution holds from the k largest observations onwards. That is, extrapolate along the line with equation

$$y = \log X_{n-k,n} + H_{k,n} \left(x + \log \frac{k+1}{n+1} \right) \quad (4.12)$$

fixed at $(\log X_{n-k,n}; \log \frac{n+1}{k+1})$.

Note that in (4.12), $x = -\log \frac{k+1}{n+1}$ gives $y = \log X_{n-k,n}$ which estimates

$$\log Q\left(\frac{n-k}{n}\right) = \log Q\left(1 - \frac{k}{n}\right) \approx \log Q\left(1 - \frac{k+1}{n+1}\right).$$

To obtain an estimator for $Q(1-p)$, take $x = -\log p$ in (4.12):

$$\begin{aligned} \log \hat{Q}(1-p) &= \log X_{n-k,n} + H_{k,n} \left(\log \frac{k+1}{p(n+1)} \right) \\ \Rightarrow \hat{q}_{k,p}^+ &= \hat{Q}(1-p) = \exp\left(\log X_{n-k,n} + H_{k,n} \left(\log \frac{k+1}{p(n+1)} \right)\right) \\ &= X_{n-k,n} \left(\frac{k+1}{p(n+1)} \right)^{H_{k,n}}. \end{aligned}$$

■

The Weissman estimator has the following asymptotic characteristics. The details of the arguments are explained in Beirlant *et al.* (2004).

Result 4.5.2: Asymptotic characteristics of the first order Weissman estimator

1. The estimator is biased:

$$E\left[\log \frac{q_{k,p}^+}{Q(1-p)}\right] \approx (H_{k,n} - \gamma) \log \frac{k+1}{(n+1)p} - b_{n,k} \frac{1 - \left(\frac{(n+1)p}{k+1}\right)^\beta}{\beta}.$$

(Here β and $b_{n,k}$ are second order parameters.)

2. The asymptotic variance is:

$$AVar(\log q_{k,p}^+) = \frac{\gamma^2}{k} \left(1 + \left(\log \frac{k+1}{(n+1)p}\right)^2\right).$$

3. The distribution is approximately Normal:

$$\sqrt{k} \left(1 + \left(\log \frac{k+1}{(n+1)p}\right)^2\right)^{\frac{1}{2}} \left(\frac{q_{k,p}^+}{Q(1-p)} - 1\right) \xrightarrow{D} N(0, \gamma^2).$$

4. The asymptotic $(1-\alpha)$ confidence interval for $Q(1-p)$ is:

$$\left(\frac{q_{k,p}^+}{1 + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \frac{H_{k,n}}{\sqrt{k}} \sqrt{1 + \left(\log \frac{k+1}{(n+1)p}\right)^2}}, \frac{q_{k,p}^+}{1 - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \frac{H_{k,n}}{\sqrt{k}} \sqrt{1 + \left(\log \frac{k+1}{(n+1)p}\right)^2}} \right).$$

■

4.5.1.2 The probability approach estimator

This approach assumes a Generalised Pareto distribution (GPD) (definition 4.3.1) for the exceedances (as defined in (4.3)). The simplest assumption is to assume a strict GPD. Simple inversion of the GPD then yields

$$Q(1-p) = U\left(\frac{1}{p}\right) = \frac{\sigma}{\gamma}(p^{-\gamma} - 1), \quad \gamma \neq 0.$$

Any estimates for the parameters may be substituted in the above equation (e.g. maximum likelihood estimates, method of moments estimates, or probability weighted moments estimates). Note that since the underlying distribution is in the Fréchet class, the derivations in this section will only be done for cases including $\gamma > 0$.

The assumption can be relaxed to assume that the exceedances follow an *approximate* GPD. The first step is then to find the survival function for the original, underlying distribution denoted by $\bar{F}(x) = P(X > x)$ below (recall that the exceedances are defined to be $\{Y_j = X_i - u \text{ given } X_i > u\}_{j=1}^{Nu}$).

Let $\bar{H}(\cdot)$ denote the GPD survival function. Then the probability approach assumes that

$$\begin{aligned} P(X - t > y | X > t) &= \frac{\bar{F}(t+y)}{\bar{F}(t)} \approx \bar{H}(y) \\ \Rightarrow \bar{F}(t+y) &\approx \bar{F}(t)\bar{H}(y) \\ \Rightarrow \bar{F}(x) &\approx \bar{F}(t)\bar{H}(x-t), \quad x = t+y. \end{aligned}$$

Substituting the GPD survival function, this yields the result

$$\bar{F}(x) \approx \bar{F}(t) \left(1 + \gamma \frac{(x-t)}{\sigma}\right)^{-\frac{1}{\gamma}}. \tag{4.13}$$

The above expression can now be used to prove the following result:

Result 4.5.3: Probability view estimator for extreme quantiles

$$\hat{U}\left(\frac{1}{p}\right) = \hat{Q}(1-p) = X_{n-k,n} + \frac{\hat{\sigma}}{\hat{\gamma}} \left[\left(\frac{k}{pn}\right)^{\hat{\gamma}} - 1 \right]. \quad \blacksquare$$

Proof:

$$\begin{aligned}
 p &= \bar{F}(x) \\
 \Rightarrow p &= \bar{F}(t) \left(1 + \gamma \frac{(x-t)}{\sigma}\right)^{-\frac{1}{\gamma}} \\
 \Rightarrow \left(\frac{p}{\bar{F}(t)}\right)^{-\gamma} &= \left(1 + \gamma \frac{(x-t)}{\sigma}\right) \\
 \Rightarrow \gamma \frac{(x-t)}{\sigma} &= \left(\frac{p}{\bar{F}(t)}\right)^{-\gamma} - 1 \\
 \Rightarrow x &= \bar{F}^{-1}(p) = Q(1-p) = U\left(\frac{1}{p}\right) = t + \frac{\sigma}{\gamma} \left[\left(\frac{p}{\bar{F}(t)}\right)^{-\gamma} - 1\right].
 \end{aligned}$$

Now set $t = X_{n-k,n}$, $\bar{F}(t) \approx \frac{k}{n}$ and replace σ and γ by a relevant estimate. ■

The delta method implies that

$$\sqrt{N_u} \left(\hat{U}\left(\frac{1}{p}\right) - U\left(\frac{1}{p}\right) \right) \xrightarrow{D} N(0, \xi^T V \xi)$$

where $V = V_1, V_2$ or V_3 depending on whether the maximum likelihood, method of moments, or method of probability weighted moments is used (refer Beirlant *et al.*, 2004 for details) and

$$\xi^T = \left[\frac{\partial U\left(\frac{1}{p}\right)}{\partial \gamma}, \frac{\partial U\left(\frac{1}{p}\right)}{\partial \sigma} \right] = \left[-\frac{\sigma}{\gamma^2} (p^{-\gamma} - 1) - \frac{\sigma}{\gamma} p^{-\gamma} \log p, \frac{1}{\gamma} (p^{-\gamma} - 1) \right].$$

An approximate $100(1-\alpha)\%$ confidence interval for $U\left(\frac{1}{p}\right)$ is then given by

$$\hat{U}\left(\frac{1}{p}\right) \pm \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \sqrt{\frac{\hat{\xi}^T \hat{V} \hat{\xi}}{N_u}}.$$

Alternatively, profile likelihood confidence intervals can be constructed using the profile likelihood ratio statistic

$$L^*\left(U\left(\frac{1}{p}\right)\right) = \max_{\gamma|U\left(\frac{1}{p}\right)} L\left(\gamma, U\left(\frac{1}{p}\right)\right).$$

The $(1-\alpha)100\%$ profile log likelihood confidence interval that corresponds to the testing of the hypothesis

$$H_0 : U\left(\frac{1}{p}\right) = U_0\left(\frac{1}{p}\right),$$

is then given by

$$\left\{U_0\left(\frac{1}{p}\right) : -2 \log\left(\frac{L^*\left(U_0\left(\frac{1}{p}\right)\right)}{\max_{U\left(\frac{1}{p}\right)} L^*\left(U\left(\frac{1}{p}\right)\right)}\right) \leq \chi_1^2(1-\alpha)\right\}.$$

4.5.2 Second order refinements

The Weissman estimator $\hat{q}_{k,p}^+$ was derived by assuming a strict Pareto distribution for large enough observations. This approach therefore ignores the slowly varying part of the Pareto-type distribution. As a result, the bias of $\hat{q}_{k,p}^+$ may potentially be very large. Second order refinements aim to reduce this bias by making use of second order information. There are two views that yield second order refinements: the quantile view and the probability view. The quantile view is discussed next. The results in the subsequent sections follow from Beirlant *et al.* (2004).

4.5.2.1 Quantile view

Second order parameters need to be introduced. One way to understand this, is to consider Renyi's exponential representation in terms of log-spacings. The distribution of these log-spacings can then allow for second order parameters.

Definition 4.5.1: Renyi's log spacings

If $X_{i,n}$ denotes the i^{th} order statistic of a sample of size n , then Renyi's log spacings are defined as

$$Z_j = j(\log X_{n-j+1,n} - \log X_{n-j,n}), \quad j=1,2,\dots,n. \quad \blacksquare$$

Note that the Hill estimator is

$$H_{k,n} = \bar{Z}_k = \frac{1}{k} \sum_{j=1}^k Z_j = \frac{1}{k} \sum_{j=1}^k (\log X_{n-j+1,n} - \log X_{n-j,n}).$$

As a first order approximation, write $Z_j = \gamma E_j$, where $E_j \sim \exp(1)$. This implies that $Z_j \sim \exp(\frac{1}{\gamma})$. However, the second order approximation of Z_j is

$$Z_j \approx (\gamma + (\frac{j}{k+1})^\beta b_{n,k}) E_j, j = 1, 2, \dots, k. \quad (4.14)$$

As a first attempt to improve upon the first order estimate $\hat{q}_{k,p}^+$, define

$$\hat{q}_{k,p}^{(0)} = X_{n-k,n} \left(\frac{k+1}{p(n+1)} \right)^{\gamma_{ML}^+} \quad (4.15)$$

Note that this is the same as the first order estimate, except that the Hill estimate $H_{k,n}$ is replaced by the maximum likelihood estimate γ_{ML}^+ of γ . This estimate is obtained from maximum likelihood methods applied to (4.14). The bias of $\hat{q}_{k,p}^{(0)}$ is lower than that of $\hat{q}_{k,p}^+$. However, the bias can be further improved upon by including a correction term, yielding the following estimate:

Result 4.5.4: Second order estimate of extreme quantile

Taking $\hat{\gamma}_{ML}^+$, $\hat{\beta}$ and $\hat{b}_{n,k}$ as the maximum likelihood estimates based on the approximation of (4.14),

$$\hat{q}_{k,p}^{(1)} = X_{n-k,n} \left(\frac{k+1}{p(n+1)} \right)^{\gamma_{ML}^+} \exp(\hat{b}_{n,k} \left(\frac{1 - (\frac{k+1}{(n+1)p})^{\hat{\beta}}}{\hat{\beta}} \right)). \quad (4.16) \blacksquare$$

Proof:

This proof considers $\frac{Q(1-p)}{X_{n-k,n}}$ and uses second order information about the slowly varying function ℓ_u .

Firstly, it is noted that to a second order approximation (Beirlant *et al.*, 2004)

$$\frac{\ell_u(xu)}{\ell_u(x)} \approx \exp(b(x)h_{-\beta}(u)),$$

where

$$h_{-\beta}(u) = \frac{u^{-\beta} - 1}{-\beta} \text{ and } b(y) = a(U^{-1}(y)).$$

Secondly, write

$$\begin{aligned} Q(1-p) &= U\left(\frac{1}{p}\right) \\ &= \left(\frac{1}{p}\right)^\gamma \ell_u\left(\frac{1}{p}\right) \end{aligned}$$

and, with $U_{j,n}$ denoting the j^{th} order statistic of a Uniform(0,1) sample of size n ,

$$\begin{aligned} X_{n-k,n} &= Q(U_{n-k,n}) \\ &= Q(1-U_{k+1,n}) \\ &= U(U_{k+1,n}^{-1}) \\ &= U_{k+1,n}^{-\gamma} \ell_u(U_{k+1,n}^{-1}), \end{aligned}$$

then

$$\begin{aligned} \frac{Q(1-p)}{X_{n-k,n}} &= \frac{p^{-\gamma} \ell_u\left(\frac{1}{p}\right)}{U_{k+1,n}^{-\gamma} \ell_u(U_{k+1,n}^{-1})} \\ &\approx \left(\frac{U_{k+1,n}}{p}\right)^\gamma \exp\left\{b(U_{k+1,n}^{-1})\left[\frac{1 - \left(\frac{U_{k+1,n}}{p}\right)^{-\beta}}{\beta}\right]\right\} \\ &\approx \left(\frac{k+1}{(n+1)p}\right)^\gamma \exp\left\{b\left(\frac{n+1}{k+1}\right)\left[\frac{1 - \left(\frac{k+1}{(n+1)p}\right)^{-\beta}}{\beta}\right]\right\}, \end{aligned}$$

with the approximation $U_{k+1,n} \approx \frac{k+1}{n+1}$ since $E[U_{k+1,n}] = \frac{k+1}{n+1}$.

$$\Rightarrow Q(1-p) \approx X_{n-k,n} \left(\frac{k+1}{(n+1)p}\right)^\gamma \exp\left\{b\left(\frac{n+1}{k+1}\right)\left[\frac{1 - \left(\frac{k+1}{(n+1)p}\right)^{-\beta}}{\beta}\right]\right\}.$$

Replacing the parameters in the right-hand side of the above equation by their maximum likelihood estimators, (4.16) follows. ■

4.5.2.2 Probability view

The second order refinement of the probability view fits a Perturbed Pareto Distribution (PPD) to the relative excesses above a certain threshold. The reasoning for this is firstly explained, followed by the resultant PPD.

A relative exceedance, given it is above a certain threshold, is defined as $Y = \frac{X}{t} \mid X > t$. If

$t = X_{n-k,n}$ then $Y_j \square \frac{X_{n-j+1,n}}{X_{n-k,n}}$, $j = 1, 2, \dots, k$. Previously from (4.3) it was assumed that the

relative exceedances follow a strict Pareto distribution. This was the first order approximation that ignored the slowly varying component. Second order information about the slowly varying component will now be used. For this the following result from Beirlant *et al.* (2004) will be used:

$$\frac{\ell(xu) - \ell(x)}{a_2(x)\ell(x)} \approx h_{-\tau}(u) = \frac{u^{-\tau} - 1}{-\tau} \text{ for } x \text{ large enough.} \quad (4.17)$$

In (4.17), $a_2(x) \rightarrow 0, x \rightarrow \infty$ and is regularly varying of order $-\tau$ ($\tau > 0$). Rewrite (4.17) slightly so that it is in a form that will be useful later:

$$\begin{aligned} \frac{\ell(xu) - \ell(x)}{\ell(x)} &\approx a_2(x)h_{-\tau}(u) \\ \Rightarrow \frac{\ell(xu)}{\ell(x)} &\approx 1 + a_2(x)h_{-\tau}(u) \end{aligned} \quad (4.18)$$

Now, since F is Pareto-type, and using (4.18) it follows that

$$\begin{aligned}
 P(X > tx | X > t) &= \frac{1 - F(tx)}{1 - F(t)} \\
 &= \frac{(tx)^{-\frac{1}{\gamma}} \ell(tx)}{(t)^{-\frac{1}{\gamma}} \ell(t)} \\
 &= x^{-\frac{1}{\gamma}} \frac{\ell(tx)}{\ell(t)} \\
 &\approx x^{-\frac{1}{\gamma}} [1 + a_2(t) h_{-\tau}(x)] \\
 &= x^{-\frac{1}{\gamma}} [1 - a_2(t) \tau^{-1} (x^{-\tau} - 1)].
 \end{aligned}$$

Therefore the survival function of the relative exceedances is that of a PPD. The findings are summarised in the result below.

Result 4.5.5: Perturbed Pareto Distribution for relative exceedances

For a high enough threshold t , the relative exceedances follow a PPD distribution

$$1 - F_t(x) = P\left(\frac{X}{t} > x | X > t\right) = (1 - c)x^{-\frac{1}{\gamma}} + cx^{-\frac{1}{\gamma} - \tau}$$

with $c = -a_2(t)\tau^{-1}$, $c \in (-\frac{1}{\tau\gamma}, 1)$, $\tau > 0$, $\gamma > 0$ and $x > 1$. ■

The parameters of the above distribution can be solved iteratively using maximum likelihood. This is no easy task, however, since the likelihood is very “flat”. To help with this task, consider the work done by Berning (2013). For mathematical convenience, adjust the notation slightly so that the survival function of the PPD is given by

$$\bar{F}_{PPD}(x) = (1 - c)x^{-1/\gamma} + cx^{-(1-\rho)/\gamma},$$

where

$$x \geq 1, \gamma > 0, \rho < 0 \text{ and } \rho^{-1} \leq c \leq 1.$$

The log likelihood to be maximised is

$$\log L(y_i; \rho, c, \gamma) = -k \log \gamma - \left(\frac{1}{\gamma} + 1\right) \left[\sum_{i=1}^k \log(y_i) \right] + \sum_{i=1}^k \log[1 - c + c(1 - \rho)y_i]$$

with k the number of exceedances above the threshold and $Y_i = \frac{X_{n-i+1,n}}{X_{n-k,n}}$.

Since the likelihood is very “flat”, it is advised to use an external estimator for ρ and to further restrict the ranges of the parameters c and γ .

The external estimator, $\hat{\rho}_3$, as proposed by Gomes and Martins (2002) is:

Definition 4.5.2: Gomes and Martins (2002) external estimator for ρ

$$\hat{\rho}_3 = - \left| \frac{3(T_n^{(0)}(k_1) - 1)}{(T_n^{(0)}(k_1) - 3)} \right|$$

with

$$T_n^{(0)}(k) = \frac{\ln(M_n^{(1)}(k)) - \frac{1}{2}\ln(M_n^{(2)}(k)/2)}{\frac{1}{2}\ln\left(\frac{M_n^{(2)}(k)}{2}\right) - \frac{1}{3}\ln(M_n^{(2)}(k)/6)},$$

$$M_n^j(k) = \frac{1}{k} \sum_{i=1}^k \left[\ln\left(\frac{X_{n-i+1,n}}{X_{n-k,n}}\right) \right]^j$$

and

$$k_1 = \min(n - 1, \lfloor \frac{2n}{\ln(\ln(n))} \rfloor). \quad \blacksquare$$

From extensive simulations, it was found that applying the restriction $c \leq 0.5$, substantially improves the EVI estimation accuracy (Berning, 2013). For $0 \leq c \leq 1$ the PPD survivor function in result 4.5.5 is the weighted sum of a first order component $x^{-1/\gamma}$, and a second order component $x^{-(1-\rho)/\gamma}$. By restricting c to be less than or equal to 0.5, we ensure that there is more weight on the first order component.

According to Berning (2013), empirical studies showed that the estimate of γ almost never exceeds 1.5. For the purpose of this analysis, it is sufficient to work with an upper limit of 1, i.e. we have $0 < \gamma < 1$ in the maximum likelihood estimation procedure.

The second order refinement to the probability approach estimate of the extreme quantile is the estimated $(1-p)^{\text{th}}$ quantile of the PPD distribution (p small), with the parameter estimates obtained as described above. This quantile will have to be solved numerically, since no explicit solution exists.

4.6 Derivation of VaR and ES using the POT approach

This section deals with the application of EVT to determine VaR and ES. VaR and ES are related to the quantiles of the underlying return distribution, which is assumed to be a fat-tailed EVT distribution (a Pareto-type distribution). Following the convention in the literature, the POT approach to EVT will be used. Firstly, however, the return for EVT models needs to be re-defined:

Definition 4.6.1: Return for EVT models

$$X_t = -R_t = (\ln P_{t-1} - \ln P_t) \approx \frac{P_{t-1} - P_t}{P_{t-1}}. \quad \blacksquare$$

The returns of definition 3.1.1 is therefore multiplied by (-1) to change the extreme negative returns into positive values. This enables the application of Extreme Value theory to the far right tail of the distribution and simplifies the mathematics.

Next, the results for VaR and ES are proven. The approach for the derivation of the VaR is along the lines of the approach followed by McNeil (1999). The ES proof is based on the work of Fernandez (2003).

The following notation is used in the proof of results 4.6.1 and 4.6.2:

F_X : The distribution function of the daily percentage returns X_t .

u : The threshold level of returns. Only the returns $\{X_t > u\}$ will be considered.

n : The total number of returns.

N_u : The number of returns above the threshold u .

F_U : The conditional distribution of the absolute exceedances.

$$F_u(y) = P(X - u < y | X > u) = \frac{F_X(u+y) - F_X(u)}{1 - F_X(u)}. \quad (4.19)$$

Result 4.6.1: VaR using the POT approach

$$VaR_\alpha = x_{1-\alpha} \approx u + \frac{\sigma}{\gamma} \left\{ \left[\frac{n}{N_u} (\alpha) \right]^{-\gamma} - 1 \right\} . \quad \blacksquare$$

Proof:

F_U can be approximated by a GPD if u is sufficiently large (Beirlant *et al.*, 2004; Pickands, 1975; Balkema and de Haan, 1974).

$F_X(u)$ can be approximated by the proportion of the observed returns below the threshold:

$$F_X(u) \approx \frac{n - N_u}{n} .$$

Equation (4.19) can be rewritten to obtain an expression for F_X (also replace $u + y$ by x):

$$\begin{aligned} F_X(u + y) &= (1 - F_X(u))F_U(y) + F_X(u) \\ F_X(x) &\approx \left(1 - \frac{n - N_u}{n} \right) \left[1 - \left(1 + \frac{\gamma(x-u)}{\sigma} \right)^{\frac{-1}{\gamma}} \right] + \frac{n - N_u}{n} \\ &= \frac{N_u}{n} \left[1 - \left(1 + \frac{\gamma(x-u)}{\sigma} \right)^{\frac{-1}{\gamma}} \right] + 1 - \frac{N_u}{n} , \end{aligned}$$

yielding finally

$$F_X(x) \approx 1 - \frac{N_u}{n} \left(1 + \frac{\gamma(x-u)}{\sigma} \right)^{\frac{-1}{\gamma}} . \quad (4.20)$$

Equation (4.20) is used to obtain an approximate expression for the Value at Risk, since this is the $(1 - \alpha)^{\text{th}}$ percentile of the above distribution.

Put $F_X(x_{1-\alpha}) = 1 - \alpha$ and invert (4.20) to obtain

$$VaR_\alpha = x_{1-\alpha} \approx u + \frac{\sigma}{\gamma} \left\{ \left[\frac{n}{N_u} (\alpha) \right]^{-\gamma} - 1 \right\} . \quad \blacksquare$$

Therefore, by approximating the conditional distribution of the exceedances by a Generalised Pareto distribution, according to the POT approach of EVT, it is possible to

obtain an EVT estimation of the VaR. The unknown parameters of the GPD can be approximated by maximum likelihood.

The derivation for the ES uses the relationship between ES and VaR. It is provided next.

Result 4.6.2: ES using the POT approach

$$ES_{\alpha} \approx \frac{VaR_{\alpha}}{1-\gamma} + \frac{\sigma-\gamma u}{1-\gamma}. \quad \blacksquare$$

Proof :

Noting that if $Y \sim GPD(\sigma, \gamma)$ then $E(Y) = \frac{\sigma}{1-\gamma}$, if $\gamma < 1$ (Coles, 2001), it follows that

$$\begin{aligned} ES_{\alpha} &= E(X|X > VaR_{\alpha}) = VaR_{\alpha} + E(X - VaR_{\alpha}|X > VaR_{\alpha}) \\ &\approx VaR_{\alpha} + \frac{\sigma + \gamma(VaR_{\alpha} - u)}{1-\gamma} \\ &= VaR_{\alpha} \left(1 + \frac{\gamma}{1-\gamma}\right) + \frac{\sigma - \gamma u}{1-\gamma} \\ &= \frac{VaR_{\alpha}}{1-\gamma} + \frac{\sigma - \gamma u}{1-\gamma}. \end{aligned} \quad \blacksquare$$

4.6.1 Confidence intervals for VaR and ES

It is useful to construct confidence intervals for unknown parameters. In this case, confidence intervals are constructed for VaR and ES. These confidence intervals will contain the true values of the parameters with a pre-specified probability and therefore provide a measure of accuracy for the estimates. Gilli *et al.* (2006) constructs so-called profile log-likelihood confidence intervals. This is the approach that will be discussed next. The result for the profile log-likelihood for VaR and the proof is provided first, followed by a similar procedure for the ES.

Result 4.6.3: $(1 - \beta)$ profile log-likelihood confidence interval for VaR_{α}

$$\{VaR_\alpha^0: -2\log \frac{L^*(VaR_\alpha^0)}{L(\hat{\tau}, \hat{\gamma})} \leq \chi_1^2(1 - \beta)\}$$

with

$L(\hat{\tau}, \hat{\gamma})$ the log-likelihood of result 4.3.2 evaluated at the maximum likelihood estimates $\hat{\tau}$ and $\hat{\gamma}$;
 $L^*(VaR_\alpha) = \max_\gamma L_0(\gamma, VaR_\alpha)$ the profile log-likelihood for VaR_α ; and,
 $\chi_1^2(1 - \beta)$ the $(1 - \beta)^{th}$ quantile of a Chi-squared distribution with one degree of freedom. ■

Proof:

Consider the GPD of definition 4.3.1,

$$H_{\gamma, \sigma}(y) = 1 - (1 + \frac{\gamma y}{\sigma})^{-\frac{1}{\gamma}}, y \in (0, \infty), \gamma > 0,$$

write σ as a function of VaR_α and γ :

$$\begin{aligned} VaR_\alpha &\approx u + \frac{\sigma}{\gamma} \{[\frac{n}{N_u}(\alpha)]^{-\gamma} - 1\} \\ \Rightarrow \gamma (VaR_\alpha - u) &= \sigma([\frac{n}{N_u}(\alpha)]^{-\gamma} - 1) \\ \Rightarrow \sigma &= \frac{\gamma(VaR_\alpha - u)}{[\frac{n}{N_u}(\alpha)]^{-\gamma} - 1}. \end{aligned}$$

Substituting this into the $H_{\gamma, \sigma}(y)$ yields:

$$H_{\gamma, VaR_\alpha}(y) = 1 - (1 + y \frac{([\frac{n}{N_u}(\alpha)]^{-\gamma} - 1)}{VaR_\alpha - u})^{-\frac{1}{\gamma}}.$$

Differentiating the above, the density function is obtained:

$$h_{\gamma, VaR_\alpha}(y) = \frac{([\frac{n}{N_u}(\alpha)]^{-\gamma} - 1)}{\gamma(VaR_\alpha - u)} (1 + y \frac{([\frac{n}{N_u}(\alpha)]^{-\gamma} - 1)}{VaR_\alpha - u})^{-\frac{1}{\gamma} - 1}.$$

Let

$$L_0(\gamma, VaR_\alpha) = \sum_{i=1}^{N_u} \log(h_{\gamma, VaR_\alpha}(y_i)),$$

where $\{y_i\}$ are the exceedances as defined in (4.3).

Hence the profile log-likelihood for VaR_α follows as

$$L^*(VaR_\alpha) = \max_\gamma L_0(\gamma, VaR_\alpha)$$

and the confidence interval is

$$\{VaR_{\alpha}^0: -2\log \frac{L^*(VaR_{\alpha}^0)}{L(\hat{\tau}, \hat{\gamma})} \leq \chi_1^2(1 - \beta)\} . \quad \blacksquare$$

Result 4.6.4: $(1 - \beta)$ Profile log-likelihood confidence interval for ES_{α}

$$\{ES_{\alpha}^0: -2\log \frac{L^*(ES_{\alpha}^0)}{L(\hat{\tau}, \hat{\gamma})} \leq \chi_1^2(1 - \beta)\} .$$

with $L^*(ES_{\alpha}) = \max_{\gamma} L_0(\gamma, ES_{\alpha})$ the profile likelihood for VaR_{α} . ■

Proof:

The proof follows a similar process as the proof of result 4.6.3. Write σ as a function of ES_{α} and γ , and substitute into the GPD:

$$H_{\gamma, ES_{\alpha}}(y) = 1 - \left(1 + y \frac{(\gamma + [\frac{n}{Nu}(\alpha)]^{-\gamma} - 1)^{-1}}{(ES_{\alpha} - u)(1 - \gamma)}\right)^{\frac{-1}{\gamma}} .$$

The derivative of the above yields the density function:

$$h_{\gamma, ES_{\alpha}}(y) = \frac{(\gamma + [\frac{n}{Nu}(\alpha)]^{-\gamma} - 1)}{\gamma(ES_{\alpha} - u)(1 - \gamma)} \left(1 + y \frac{(\gamma + [\frac{n}{Nu}(\alpha)]^{-\gamma} - 1)^{-1}}{(ES_{\alpha} - u)(1 - \gamma)}\right)^{\frac{-1}{\gamma} - 1} .$$

Let

$$L_0(\gamma, ES_{\alpha}) = \sum_{i=1}^{Nu} \log(h_{\gamma, ES_{\alpha}}(y_i))$$

then the profile likelihood for VaR_{α} is

$$L^*(ES_{\alpha}) = \max_{\gamma} L_0(\gamma, ES_{\alpha}) .$$

and the confidence interval is

$$\{ES_{\alpha}^0: -2\log \frac{L^*(ES_{\alpha}^0)}{L(\hat{\tau}, \hat{\gamma})} \leq \chi_1^2(1 - \beta)\} . \quad \blacksquare$$

This section showed that EVT, in particular the POT approach, can be useful to obtain estimates of VaR and ES. EVT is particularly relevant since the return distributions are fat-tailed. However, EVT should not be applied blindly and the reader should be aware of the limitations of EVT. Some of these are discussed next.

4.7 Limitations of EVT

Diebold *et al.* (1998) cautions against the over-use of EVT and outlines some of the pitfalls of EVT in risk management. One of these is the choice of the cut-off point for classifying an observation as extreme. This translates into selecting the threshold for exceedences used to estimate the Generalised Pareto distribution parameters (refer section 4.4.3). The trade-off between variance and bias plays an important role here.

Other problems highlighted by Diebold *et al.* (1998) are the estimation of the extreme value index, inadequate knowledge as to the sampling properties of the estimator of the extreme value index, the validity of ignoring the slowly varying part of Pareto-type distributions and the violation of certain assumptions such as that of independent identically distributed observations, usually assumed in EVT models. The last problem can be addressed by fitting a stochastic volatility model as an interim step.

Another limitation of EVT discussed by Bensalah (2000) is the application in a multivariate environment, i.e. where the returns from a portfolio of assets as a whole are considered. It is difficult to classify a n -dimensional vector as extreme ($n > 1$), since a standard definition of order in n -dimensional spaces does not exist.

One proposed method to handle a portfolio of assets, is to estimate the marginal extreme value distribution, and hence the VaR quantiles (VaR_i), for each asset. Bensalah uses the block maxima method for this. The correlations between the series of maxima are also calculated (say r_{ij}). The VaR for a portfolio of N assets, with weight w_i invested in the i^{th} asset, then takes the form (Bensalah, 2000):

$$VaR_{portfolio} = \sqrt{\sum_{i=1}^n \sum_{j=1}^n \rho_{ij} w_i w_j VaR_i VaR_j}. \quad (4.21)$$

The problem with this approach is that the correlations between the various assets may not be fully accounted for. Extreme movements in the individual asset prices need not imply an extreme movement in the portfolio as a whole.

4.8 Scaling of VaR and ES

The VaR and ES measures of results 4.6.1 and 4.6.2 apply to one day time horizons. Typically, a ten day time horizon is considered in financial applications. A procedure is therefore needed to scale the one day VaR and ES to ten day estimates. Various approaches may be followed, some of which are described in the following sections. However, there seems to be no clear cut answer as to the correct procedure. Note that attention will only be given to the scaling of the VaR estimates in what follows. The corresponding h-day ES can be derived from the h-day VaR using the relationship in result 4.6.2.

4.8.1. The square root of time rule

The simplest approach is to use the square root of time rule. This states that the h-day VaR is equal to the one day VaR multiplied by the square root of h, i.e.

$$VaR(h) = VaR(1)\sqrt{h} .$$

However, for the square root of time rule to hold, Diebold *et al.* (1997) identified three conditions. Firstly, the portfolio must remain static over the holding period. Secondly, the returns must be i.i.d., and thirdly, the returns must be normally distributed. The returns are not necessarily independent, although this assumption is needed to apply EVT. The main problem is that the returns are not normally distributed, but have fat tails. The actual volatility over h days is likely to be smaller than predicted using the square root of time rule (Ordening *et al.*, 2002).

4.8.2. The alpha root rule

The alpha root of time rule follows from Danielsson and de Vries (2000). This solves the issue of normality, since the rule holds for the Fréchet class of distributions. However, returns are still assumed to be i.i.d..

The argument is along the following lines:

Assume that for a single-period return, X ,

$$P(|X| > x) = Cx^{-\alpha},$$

then for the h-period return it holds that

$$P(X_1 + X_2 + \dots + X_h > x) = hCx^{-\alpha}$$

and therefore

$$VaR(h) = VaR(1)h^{\frac{1}{\alpha}}$$

where $\frac{1}{\alpha}$ is the EVI.

4.8.3. Empirically established scaling laws

McNeil and Frey (1998) conducted an empirical study and established scaling laws depending on the current volatility and significance level of the VaR estimate. The study found that

$$VaR(h) = VaR(1)h^{\lambda}$$

where the value of λ is given in Table 4.8.1.

Table 4.8.1: Scaling exponents for multiple day horizons

	Significance level	
	95%	99%
Low volatility	0.65	0.65
Average volatility	0.60	0.59
High volatility	0.48	0.47

Table 4.8.1 is valid for horizons up to 50 days. Note that the square root of time rule ($\lambda = 0.5$) is applicable only to high volatility periods. Low, average and high volatilities are the 5th, 50th and 95th percentiles of the estimated historical volatilities.

For the practical application in sections 8.4 and 8.5 the following steps will be followed:

1. Calculate the historical volatilities for each of the 250 day data windows. This is the distribution of the historical volatilities.
2. Calculate the 5th, 50th and 95th percentiles of this distribution.
3. Use linear interpolation to determine what scaling factor to apply to obtain the estimate for day $t+h$, depending on where the volatility observed for day t falls relative to the percentiles.

Note that instead of calculating the volatility, which is the annualised standard deviation (the standard deviation multiplied by a factor equal to the square root of 250), the standard deviations can also simply be calculated and used to determine the scaling factors. This is because the volatility is a constant multiple of the standard deviation and this will not affect the relative position of each day's volatility with respect to the entire volatility distribution.

4.9 Summary

This chapter introduced EVT. EVT is particularly useful in VaR and ES calculations, since return distributions are mostly fat-tailed, i.e. of Pareto type. The two broad approaches in EVT were discussed. These are the block maxima approach, leading to the generalised extreme value theory distribution, and the peaks over threshold approach, leading to the GPD. The latter approach is preferred, since it makes more efficient use of the data. Various estimators of the parameters of the GPD were considered.

The Fréchet class of distributions was discussed in detail, since this class includes the important Pareto-type distributions having heavy tails. The extreme value index can be estimated by the Hill estimator and is expected to be positive for Pareto-type distributions. The selection of the threshold for the exceedances was shown to be no easy task. Quantile-quantile plots, Hill plots and mean excess plots were discussed as possible selection methods.

Quantile estimation is an end goal of EVT, since VaR and ES are directly related to the quantiles of the underlying distribution. First order estimators ignore the slowly varying part of the Pareto-type distribution and may therefore lead to substantial bias. Examples of these that were considered are the Weissman estimator and the probability approach estimator. Second order refinements aim to reduce this bias by making use of second order information. Second order refinements based on the quantile view, which involves a second order correction term to the distribution of Renyi's log-spacings, as well as second order refinements to the probability view were considered. The latter led to the so-called Perturbed Pareto Distribution for relative exceedances.

The penultimate section considered how the peaks over threshold approach can be used to estimate VaR and ES measures. The results in that section are important, and will later be referred to as traditional EVT VaR and ES estimators. The chapter concluded in pointing out some limitations of EVT and options in the scaling of VaR and ES.

The next chapter combines traditional VaR and ES estimators with extreme value theory.

5. Combining EVT with stochastic volatility models

The EVT estimates for VaR and ES derived in section 4.6 of the previous chapter require that returns be independent and identically distributed. This is unlikely to be the case in practice. A method to obtain independent, identically distributed returns is therefore required as an interim step. This chapter shows that the VaR theory of chapter 3 can be used in conjunction with the EVT of chapter 4 to develop a so-called two step procedure for the estimation of VaR and ES.

5.1. Two step procedure

The two step procedure suggested by McNeil (1999) involves a combination of the GARCH conditional model with the EVT model. The idea is to first filter the returns, by applying the GARCH model, to obtain approximately independent and identically distributed residuals that can then be used in EVT. The model used in this two step procedure is given below, followed by an explanation of the two steps. The models in section 3.4 will be assumed to hold for the returns.

Definition 5.1.1: Model for the two step procedure

$$X_i = \mu_i + \sigma_i Z_i$$

where X_i is the return at time i (with historical data available for $i = 1, 2, \dots, t$), $\{Z_i\}$ are random variables distributed i.i.d., and μ_i and σ_i are the mean and standard deviation of the i^{th} return. ■

First step:

Filter the observations X_i by fitting models for the conditional variance and conditional mean similar to equations (3.9) and (3.8). The residuals from this fit have the form

$$\hat{Z}_i = \frac{X_i - \hat{\mu}_i}{\hat{\sigma}_i}$$

It is assumed (McNeil, 1999) that the residuals $\{\hat{Z}_i\}$ are approximately independent and identically distributed.

Second step:

Apply extreme value theory to these residuals. That is, \hat{Z}_i takes the place of X_t in the discussion in section 4.6. Briefly this can be summarised as follows:

1. The absolute exceedances above a certain threshold are calculated:

$$\{\hat{Z}_i - u \mid \hat{Z}_i > u\}, i=1,2,\dots, t.$$

McNeil (1999) suggests setting the threshold equal to the 90th percentile of the residuals.

2. A GPD model is fitted to these exceedances and the maximum likelihood estimates of the parameters are obtained.

The estimated distribution of the residuals is then given by (compare equation (4.20)):

$$\hat{F}_{\hat{Z}}(z) \approx 1 - \frac{N_u}{n} \left(1 + \frac{\hat{\gamma}(z-u)}{\hat{\sigma}}\right)^{-\frac{1}{\hat{\gamma}}}.$$

3. VaR is the $(1 - \alpha)^{\text{th}}$ percentile of $F_{\hat{Z}}(z)$, i.e. the estimate is:

$$\widehat{\text{VaR}}_{\alpha}(\hat{Z}) = \hat{z}_{1-\alpha} \approx u + \frac{\hat{\sigma}}{\hat{\gamma}} \left\{ \left[\frac{n}{N_u}(\alpha) \right]^{-\hat{\gamma}} - 1 \right\}. \quad (5.1)$$

Expected shortfall is the conditional expectation, estimated as

$$\widehat{\text{ES}}_{\alpha}(\hat{Z}) = E(\hat{Z} \mid \hat{Z} > \widehat{\text{VaR}}_{\alpha}(\hat{Z})) = \frac{\widehat{\text{VaR}}_{\alpha}(\hat{Z})}{1-\hat{\gamma}} + \frac{\hat{\sigma}-\hat{\gamma}u}{1-\hat{\gamma}}. \quad (5.2)$$

Equations (5.1) and (5.2) above follow from result 4.6.1 and result 4.6.2 respectively.

The two step procedure will have to be repeated each day. That is, new parameter estimates will be calculated each day. The estimates for VaR and ES of the model in definition 5.1.1 are therefore dynamic.

Results 5.1.2 and 5.1.3 follow directly from definition 5.1.1.

Result 5.1.2: Dynamic estimates of VaR for day t+1 from two step procedure

$$\widehat{\text{VaR}}_{t,\alpha} = \hat{\mu}_t + \hat{\sigma}_t \widehat{\text{VaR}}_{\alpha}(\hat{Z}) \quad \blacksquare$$

The result for the expression of the estimate of ES for day t+1, calculated on day t, is similar:

Result 5.1.3: Dynamic estimates of ES for day t+1 from the two step procedure

$$\widehat{ES}_{t,\alpha} = \hat{\mu}_t + \hat{\sigma}_t \widehat{ES}_\alpha(Z) \quad \blacksquare$$

5.2 h-day time horizons

The calculation of the VaR and ES measures for h days using 1 day returns is relatively complex. The estimates cannot simply be scaled using the so-called square-root-of-time rule, since the returns are not independent and identically distributed. Indeed, they are supposed to come from a fat-tailed GPD. As a solution, the following simulation approach is suggested by McNeil (1999).

1. Consider the residuals $Z_t = \frac{X_t - \hat{\mu}_t}{\hat{\sigma}_t}$ (obtained as described above). Fit a GPD to both tails of the residual distribution. Use an empirical estimate (historical simulation) for the residuals in the centre of the distribution and tie them together.
2. Simulate h values $Z_{t+1}, Z_{t+2}, \dots, Z_{t+h}$ independently from this composite distribution.
3. Transform these residual simulations back to simulations from the original return distribution, using the relationship $X_t = \hat{\mu}_t + \hat{\sigma}_t Z_t$ where $\hat{\mu}_t$ and $\hat{\sigma}_t$ are the estimates from the ARCH and GARCH models. This yields $X_{t+1}, X_{t+2}, \dots, X_{t+h}$.
4. The h day return is then calculated as $\sum_{i=1}^h X_{t+i}$.
5. Repeat steps 2 through 4 a large number of times (say 1000) to obtain a large number of simulated h day returns.
6. Use the returns in step 5 and apply the two-step conditional EVT procedure from the start to these. This will yield h day VaR and ES estimates.

For horizons up to 50 days, McNeil and Frey (1998) found that the h day VaR estimate is approximately equal to the 1 day VaR estimate, multiplied by a scaling factor of h^{λ_t} where λ_t depends on the significance level and volatility level. $\lambda_t \approx 0.5$, resulting in the square-root-of-time rule, only when volatility was very high.

6. Practical issues in model fitting

The practical application of the models to estimate the VaR and ES entails using a moving data window of 250 days' daily log returns. In this application, it is found that the GARCH models do not converge over certain periods and that the EVT models do not always provide sensible results. There are various reasons for this.

The first reason is that the sample size, a data window of 250 days, is very small. Classical EVT requires much larger sample sizes. A sample of size 250 provides very few observations in the tails on which the EVT models can be based. Peng *et al.* (2006) based their analysis of the Composite Index of the Shanghai Stock Exchange on a window width of 500 and later expanded it to 1000. GARCH models also need enough data for estimation. According to <http://www.portfolioprobe.com/2012/07/06/a-practical-introduction-to-garch-modeling/> less than 1000 observations are unlikely to yield informative parameter estimates.

In practice however, the VaR calculations require a window width of 250. This is because markets change rapidly and information over longer estimation periods is unlikely to be relevant in predicting the future. There is therefore a conflict between obtaining relevant, useful estimates and model convergence or performance. This study will consider, given the practical requirement of a window width of 250 days, what is the best that can be done with the available models.

The second reason why the GARCH models do not converge over certain periods is that the moving data window contains an extreme value that cannot be handled by standard optimisation procedures. In this case, EVT models can be considered as alternative models.

According to Zivot (2008), the GARCH log likelihood function is often not "well-behaved" and a global maximum cannot be achieved using standard optimisation techniques. This is especially true for more complicated GARCH models with an increased number of parameters. Starting values, optimisation algorithm choice and convergence criteria all influence the resulting estimates. The positive variance and stationarity constraints are often

forced to be ignored in practice. Therefore, GARCH parameter estimates should always be interpreted with care and awareness of their limitations is critical.

Another problem in fitting GARCH models to financial data, is that evidence shows that the estimate of α_1 in the GARCH(1,1) model is often close to zero and β_1 close to one. Indeed, α_1 is often between 0 and 0.1 and β_1 is often between 0.9 and 1 (<http://www.portfolioprobe.com/2012/07/06/a-practical-introduction-to-garch-modeling/>). In this case, the parameter β_1 is said to be “unidentified” and the maximum likelihood estimates are ill-behaved as a consequence (Zivot, 2008). The log likelihood as a function of β_1 can then exhibit multiple maxima (Ma *et al.*, 2006).

In the light of the issues highlighted in this chapter, the practical implementation of the GARCH models are considered only in areas of the data where all the GARCH models converge. The performance of the different GARCH models is compared in these areas, which are typically the lower volatility, more tranquil areas where extreme observations do not hinder the convergence process. The EVT models can theoretically be applied to the entire estimation period, although caution should be used in interpreting the estimates during less volatile periods where there are little or no extreme returns. EVT models can then be used to augment the GARCH models. That is, EVT can be used in areas where the GARCH models fail.

7. The Data

7.1. Description of the data

The log returns of the Africa Financials Index (J580) from 29/01/1996 to 30/04/2013 will be considered in the practical application. This index trades on the JSE and consists inter alia of holdings in Standard Bank Group Limited, Firstrand Limited, Old Mutual Plc, Absa Group Limited, Sanlam Limited (for a complete list, refer <http://dashboard.fin24.com/DataProducts/IndexConstituents.aspx?Ticker=J580>). Figure 7.1 plots the index values over the period 26/01/1996 to 30/04/2013. In Figure 7.2 the log returns are plotted and in Figure 7.3 the histogram is provided. All figures relating to the log returns in this section are plotted for the period 29/01/1996 to 30/04/2013. Some basic statistics pertaining to the data set are summarised in Table 7.1.

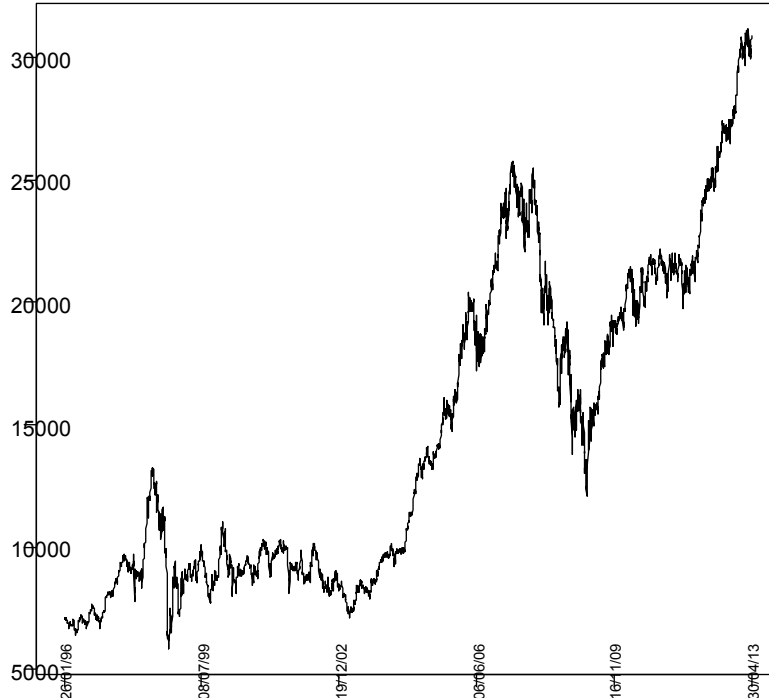


Figure 7.1: Africa Financials Index (J580) over the period 26/01/1996 to 30/04/2013.

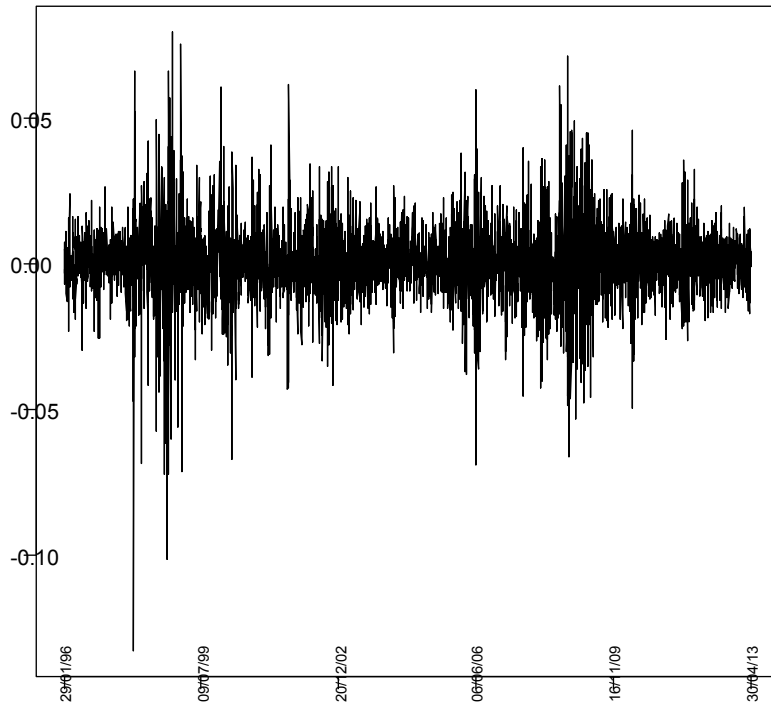


Figure 7.2: Log returns of Africa Financials Index over the period 29/01/1996 to 30/04/2013.

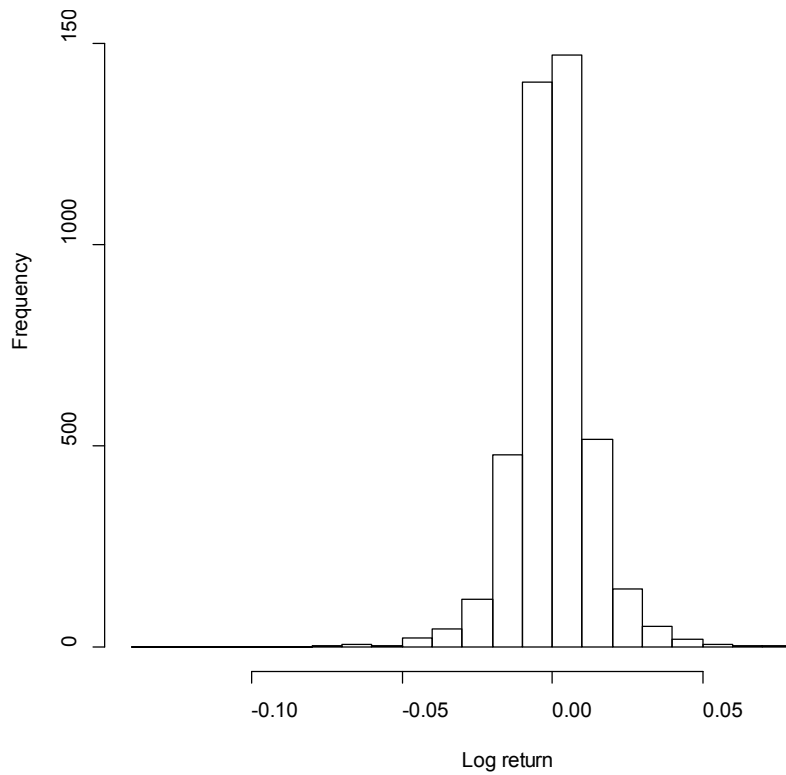


Figure 7.3: Histogram of log returns of Africa Financials Index.

Table 7.1: Descriptive statistics for the log returns of the Africa Financials Index.

Number of data points	4313
Mean	0.0003416
Median	0.0004576
Standard deviation	0.0134093
Maximum	0.07983
Minimum	-0.1331
Inter-quartile range	0.01342
Excess kurtosis	7.0253
Skewness	-0.4246
Dickey-Fuller Unit Root test statistic (lag 16)	-14.9057 (p-value<0.01)
Jarque Bera test statistic	8999.037 (p-value<2.2e-16)

The Dickey-Fuller test indicates that the series can be assumed to be stationary, since the null hypothesis of a unit root is rejected. This is to be expected, since log returns are considered. Taking the logarithm amounts to a variance stabilising transformation. In addition, first order differences of the log returns are computed, which is also a common method to achieve stationarity in a time series.

Although the series is stationary, it does not follow a normal distribution, as indicated by the large excess kurtosis (7.0253) and negative skewness (-0.4246). The negative skewness is to be expected for an index of share prices, since extreme negative returns are more likely than extreme positive returns. The Jarque Bera test strongly rejects the null hypothesis of

normality. This is in accordance with financial series being heavy tailed and exhibiting excess kurtosis.

A plot of the squared log returns (Figure 7.4) shows the volatility clustering behaviour of the time series. Low returns are followed by low returns and high returns by high returns.

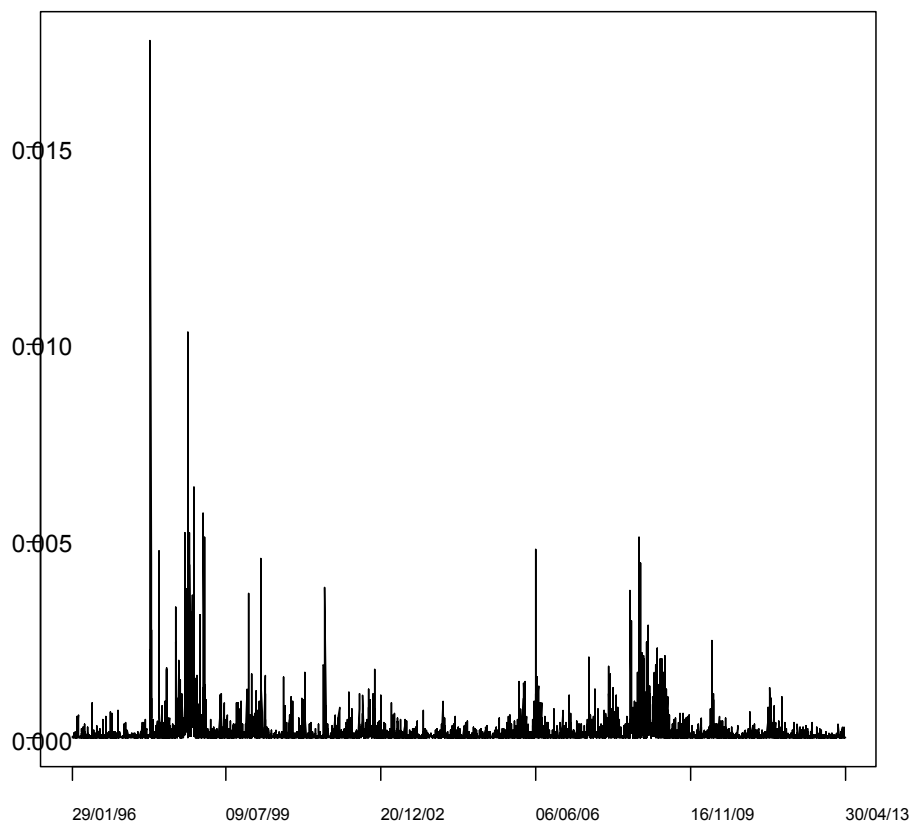
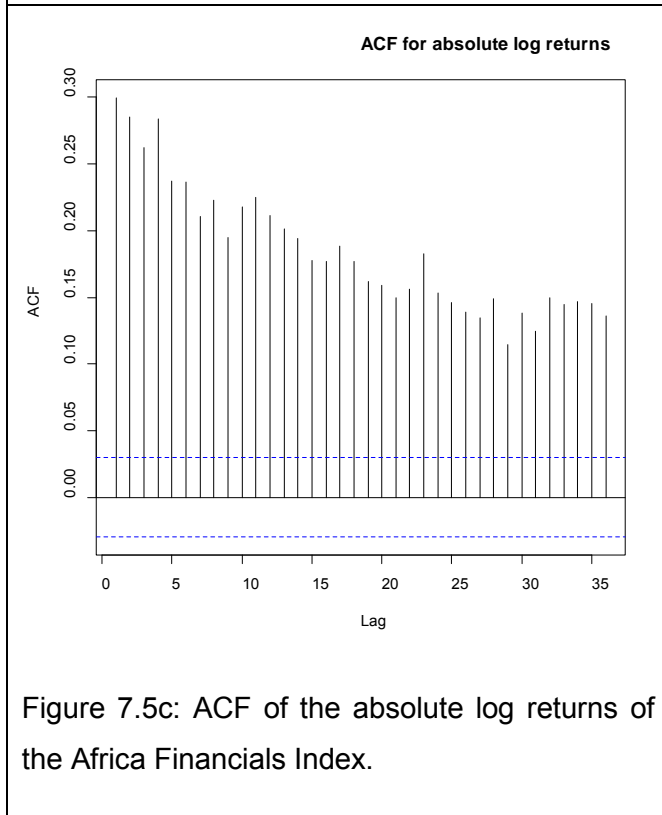
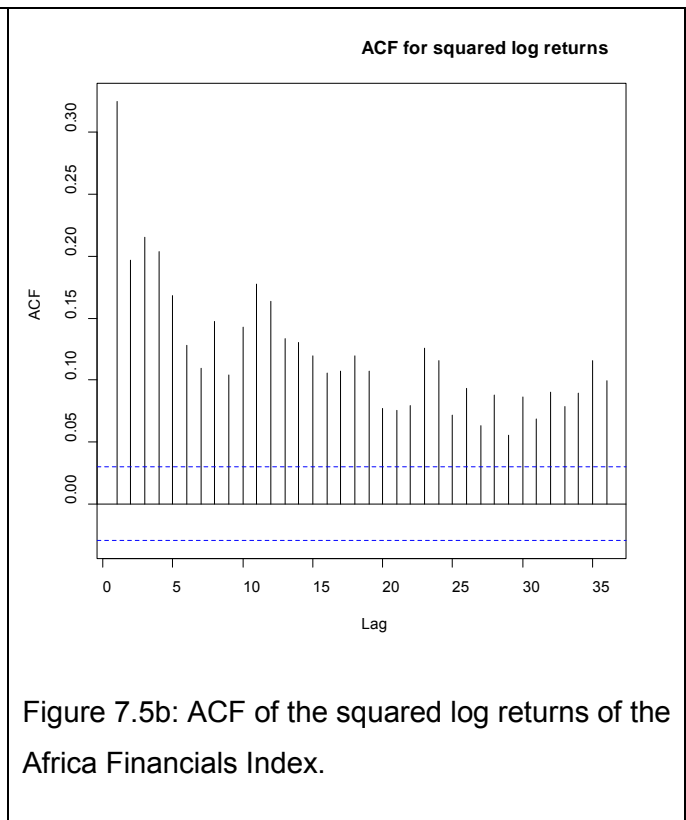
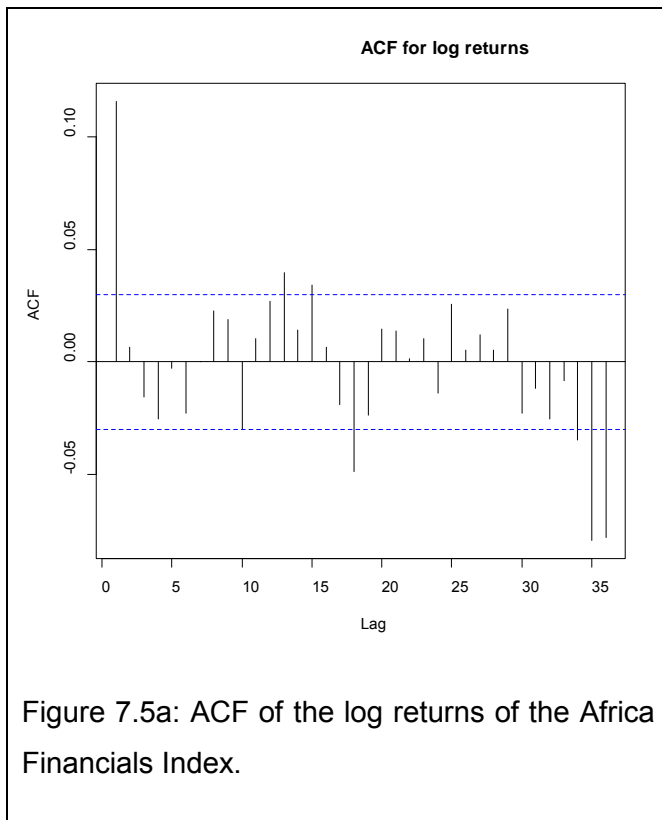


Figure 7.4: Squared log returns of the Africa Financials Index.

The plot of the autocorrelation function (ACF) of the log returns (Figure 7.5a) shows few significant autocorrelations, providing evidence of the stationarity of the series. However, the ACF of the squared returns (Figure 7.5b) and the absolute returns (Figure 7.5c) are highly significant for all lags and decay slowly. This is a result of the volatility clustering effect and provides evidence of the existence of ARCH/GARCH effects in the time series.



Another property of financial time series, is the asymmetric effect or leverage effect that the volatility is higher after negative shocks than positive shocks of the same magnitude. Zivot (2008) suggests testing for this asymmetric effect by computing the correlation between the squared log returns and that of the log return one day back. For the Africa Financials Index, this value is

$$\text{corr}(r_t^2, r_{t-1}) = -0.1282763.$$

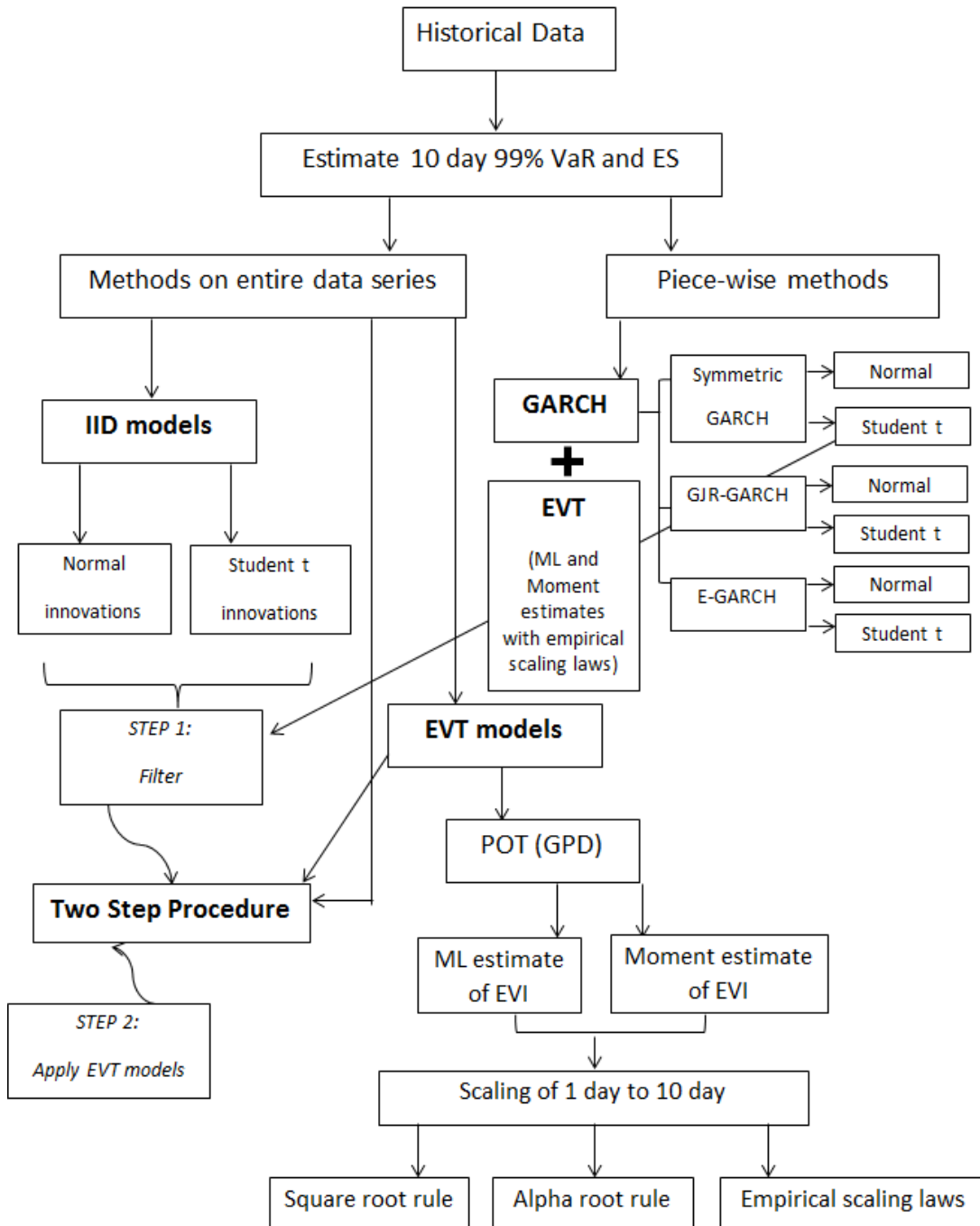
Since this value is negative, it provides evidence for leverage effects.

In summary, the log returns series is a suitable candidate for GARCH models, because of the volatility clustering effect. Asymmetric GARCH models may prove useful, because of the leverage effect. A distribution other than the normal (e.g. the student t distribution) for the errors may be considered, because of the excess kurtosis in the data. Extreme Value Theory can be applied because of the fat-tailed nature of the data. The next chapter considers the analysis of the Africa Financials data set.

8. Analysis of the Africa Financials Index Data

8.1. Schematic Overview

The following schematic overview summarises the analysis of the Africa Financials index in this chapter.



8.2. Independent identically distributed models

Two i.i.d. models for VaR and ES are considered. The normal VaR and ES follow from equations (3.4) and (3.5) of Chapter 3, whereas the student t versions follow from (3.6) and (3.7) respectively. A visual display of the analysis is provided in Figures 8.1 and 8.2. The 10 day log returns are calculated as $R_{t,10} = \log P_t - \log P_{t-10}$. The number of violations for each model is summarised in Table 8.1. These are defined as the number of times over the sample that the true returns were below the VaR or ES estimates. Note that the expected number of violations for the 99% VaR or ES measures is $0.01 \cdot 4045 = 40.45$ where 4045 is the number of days for which 10 day ahead forecasts can be calculated. Figure 8.3 graphically displays the information in a bar chart. The blue reference line is the expected number of violations.

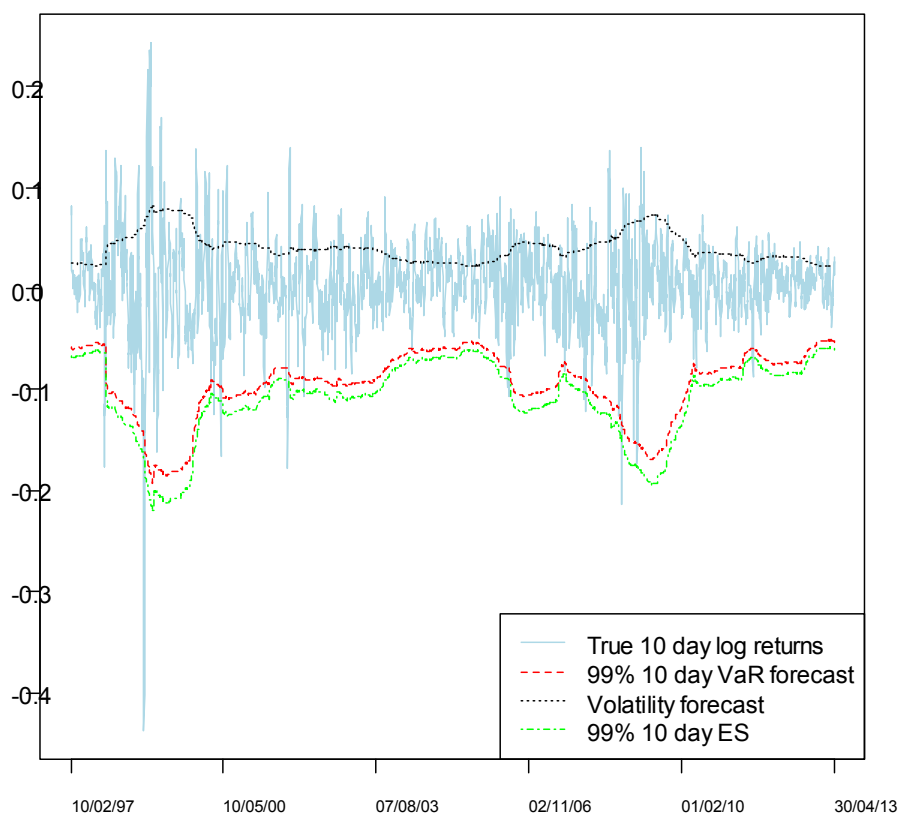


Figure 8.1: VaR and ES for the i.i.d. normal model.

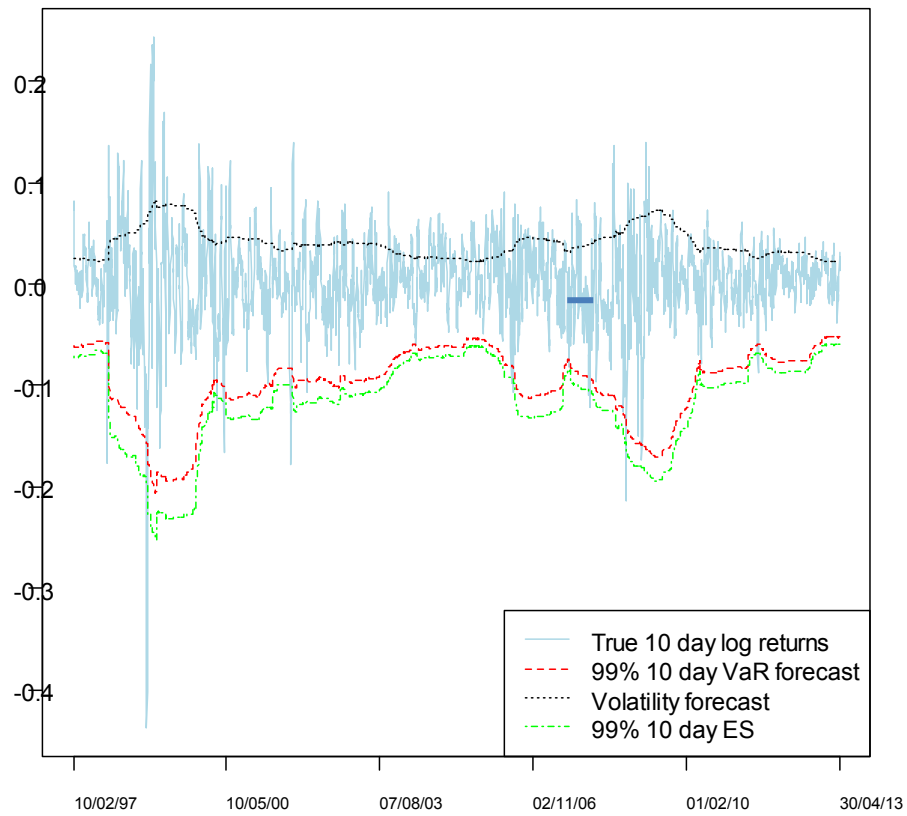


Figure 8.2: VaR and ES for the i.i.d. Student t model.

Table 8.1: Expected and observed violations of returns using a 10 day forecast horizon.

	VaR forecast	ES forecast	Expected
I.i.d. Normal	51	35	40.45
I.i.d. Student t	43	32	40.45

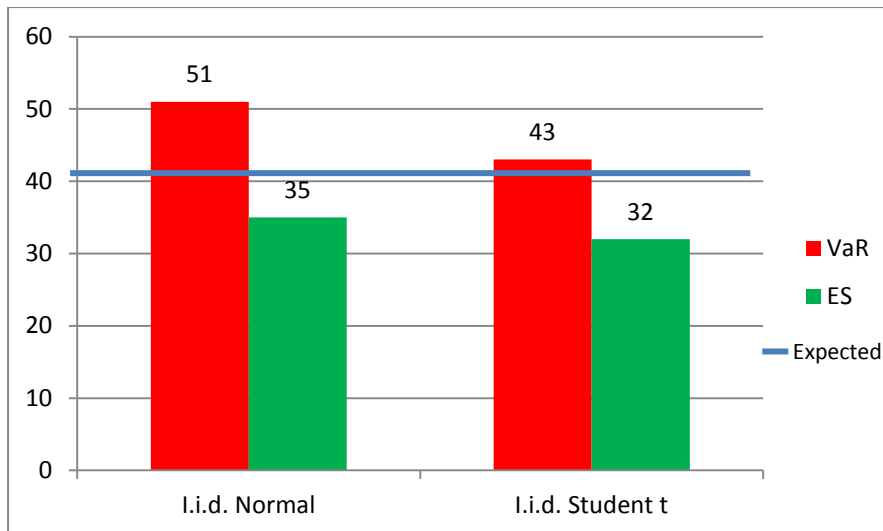


Figure 8.3: Bar chart of number of VaR and ES violations for the i.i.d normal and i.i.d. Student t models.

The VaR forecasts yield more than the expected number of violations, whereas the ES forecasts are more conservative. The model with the student t innovations seems to be superior in terms of the number of violations criterion. The application of GARCH models to the data is investigated in the next section.

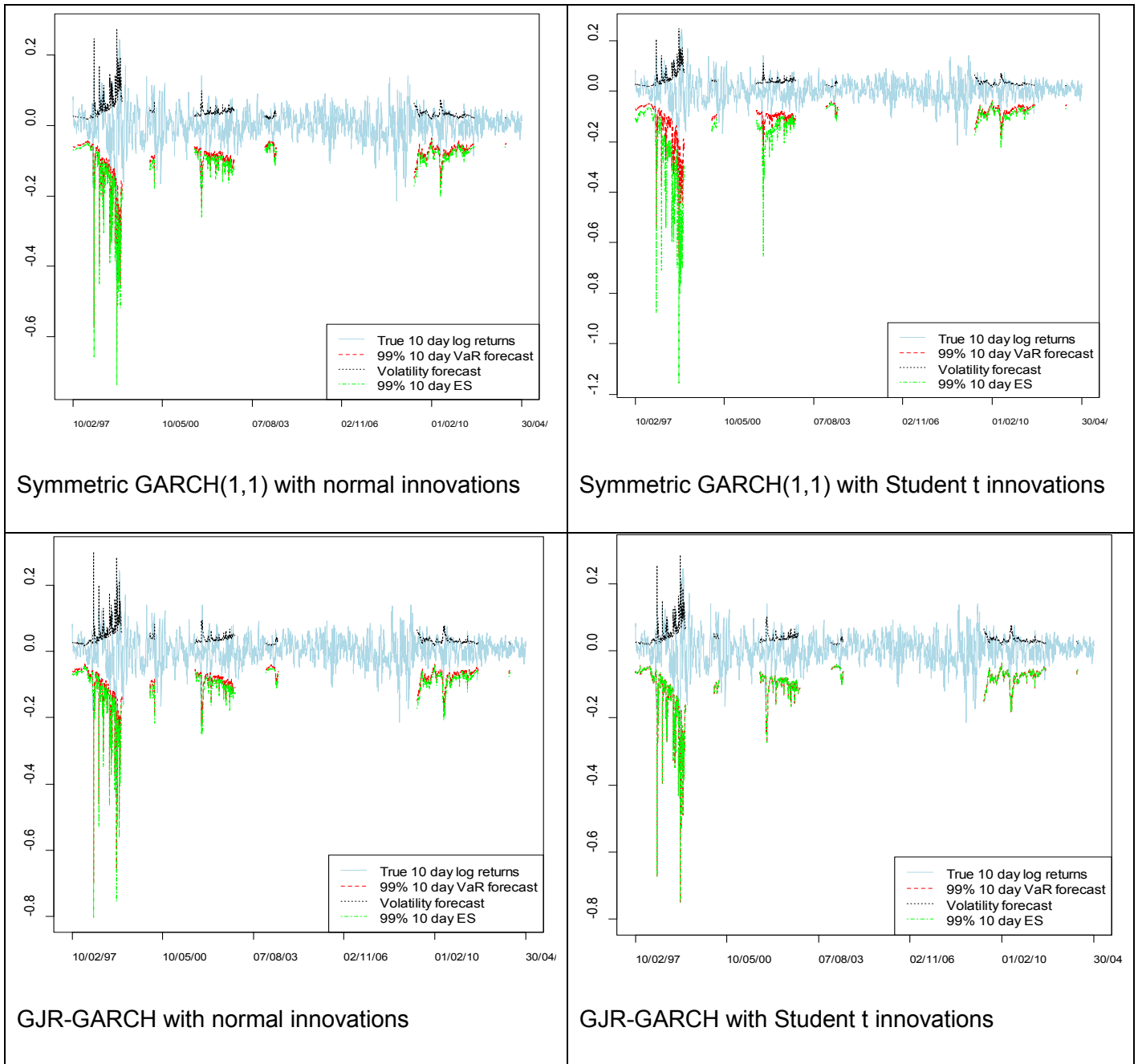
8.3. GARCH models

The GARCH models do not converge over the entire sample period. In the analysis below the GARCH models are compared over intervals where convergence is possible for all the GARCH models. The following six GARCH models are considered in this comparison (refer sections 3.4 and 3.5):

1. The symmetric GARCH model with normal innovations
2. The symmetric GARCH model with Student t innovations
3. The GJR-GARCH with normal innovations
4. The GJR-GARCH with Student t innovations
5. The E-GARCH with normal innovations
6. The E-GARCH with Student t innovations.

8.3.1. Estimation results

The results of the volatility forecast over the next ten days, the 99% ten day VaR as well as the 99% ten day ES are graphically displayed in Figure 8.4. Figure 8.5 gives the i.i.d. estimates limited to the convergence periods. Table 8.2 and Figure 8.6 summarise the expected and actual number of violations for the different estimates for the different methods so far, also including the i.i.d. models. Note that the total number of returns in the convergence periods is 1514. Therefore the expected number of violations is 15.14.



Symmetric GARCH(1,1) with normal innovations

Symmetric GARCH(1,1) with Student t innovations

GJR-GARCH with normal innovations

GJR-GARCH with Student t innovations

Figure 8.4 (continued): Estimation results for the different GARCH models applied to the Africa Financials Index over the period 10/02/1997 to 30/04/2013.

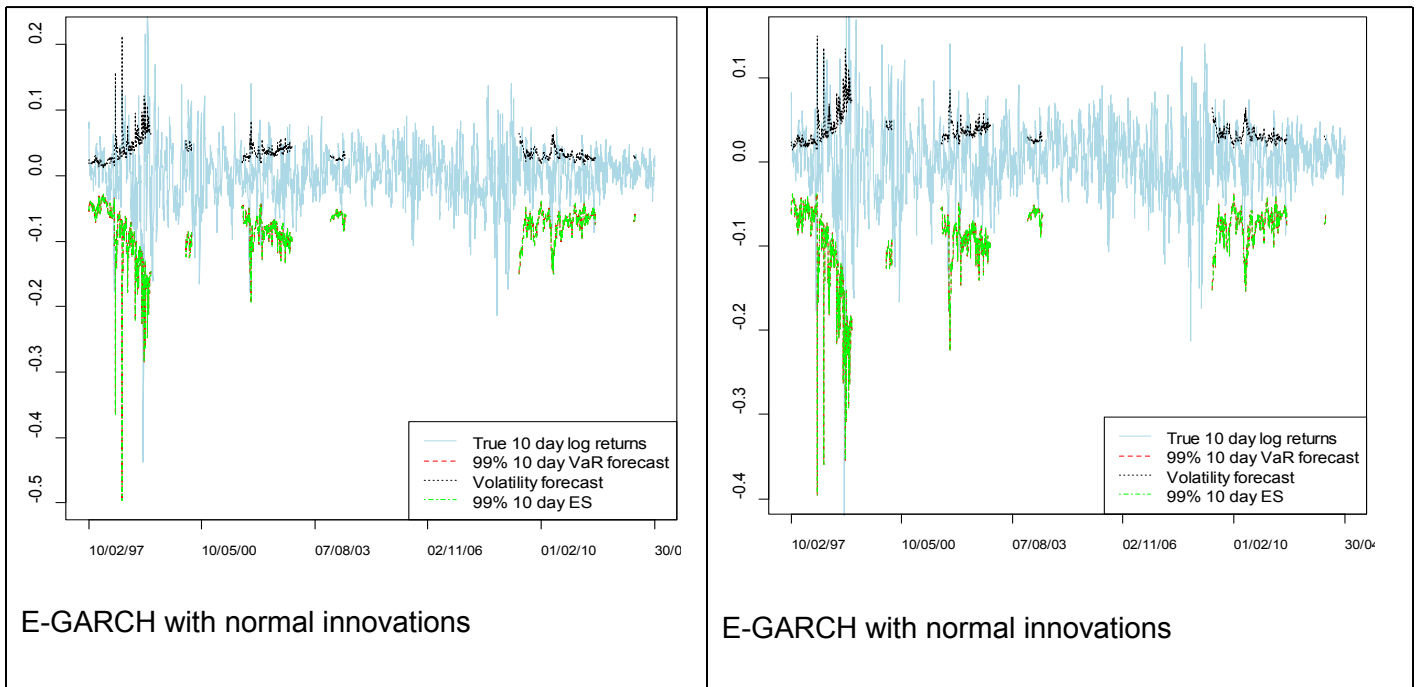


Figure 8.4 (continued): Estimation results for the different GARCH models applied to the Africa Financials Index over the period 10/02/1997 to 30/04/2013.

For comparison, the i.i.d. models can also be used over the selected periods only, where GARCH forecasting is possible. These are given in Figure 8.5.

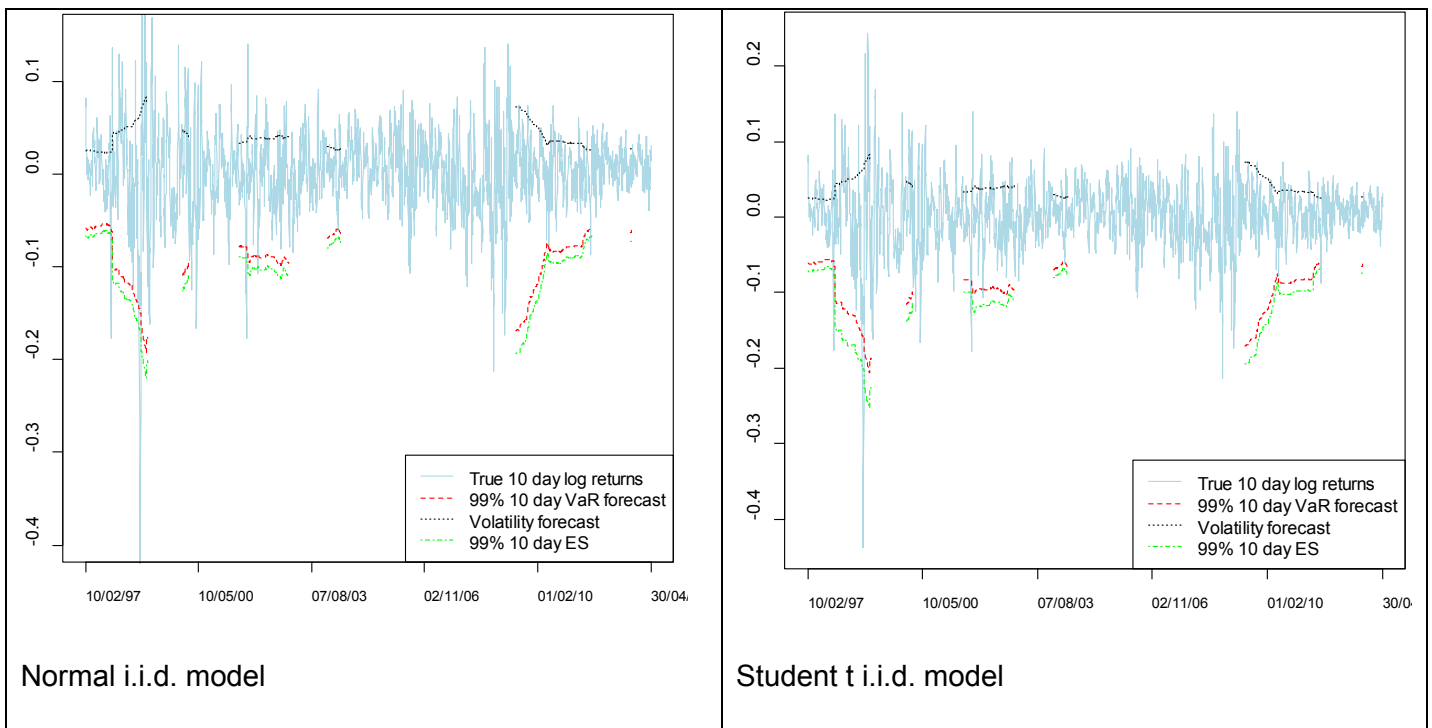


Figure 8.5: Estimation results for the i.i.d. models over the convergence period.

Table 8.2: Expected and observed violations of returns for different models over the convergence periods.

	VaR forecast	ES forecast	Expected
I.i.d Normal	31	27	15.14
I.i.d Student t	29	25	15.14
Symmetric GARCH (N)	38	28	15.14
Symmetric GARCH (t)	30	13	15.14
GJR-GARCH (N)	43	32	15.14
GJR-GARCH (t)	35	35	15.14
E-GARCH (N)	42	42	15.14
E-GARCH (t)	36	36	15.14

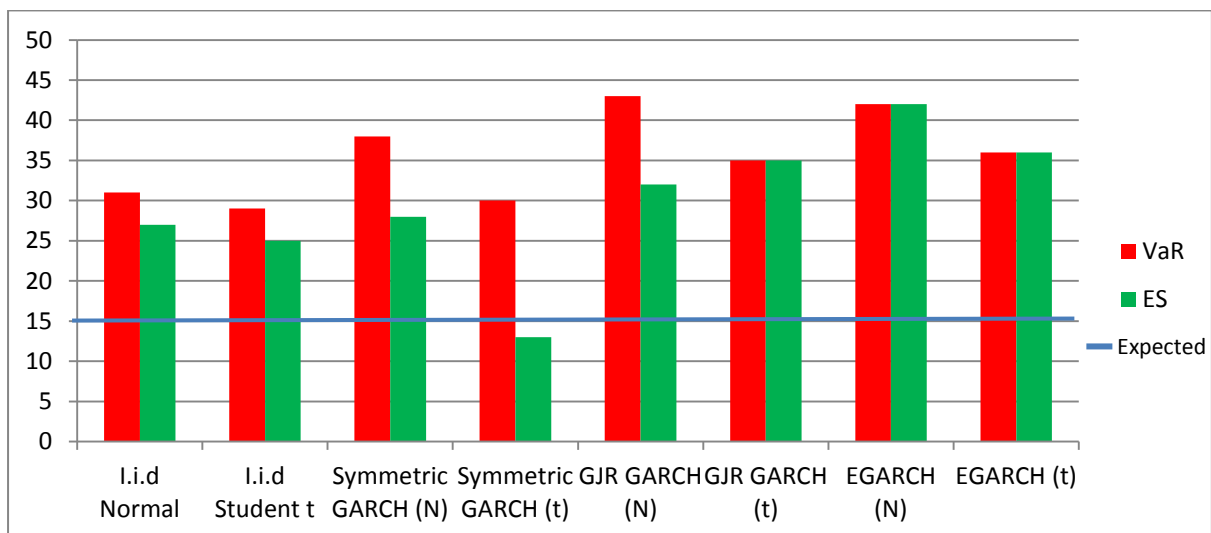


Figure 8.6: Bar chart of expected and observed violations of returns for different models over the convergence periods.

The ES produces a consistently lower number of violations than the VaR counterparts. This provides evidence for the superiority of ES over VaR in that the ES estimates are more conservative. The model producing the least number of VaR violations is the i.i.d. Student t model, whereas the model producing the least number of ES violations is the symmetric GARCH model with Student t innovations. The latter model also does well with respect to the VaR violations.

All models except the symmetric GARCH ES with Student t innovations yield more than the expected number of violations. The models therefore tend to underestimate the occurrence of extreme events.

As is to be expected, the models with Student t innovations perform better than those with normal innovations, since the Student t distribution can take better account of the fat tails of the data because of the additional kurtosis. The i.i.d. models perform remarkably well compared to the stochastic volatility models. The i.i.d Student t model seems to be a very attractive option. The symmetric GARCH model with t innovations also performs very well. The asymmetric GARCH models do not provide superior results over the symmetric models. This means that accounting for skewness and leverage effects do not lead to better results.

It must be kept in mind that this analysis was only performed for periods where the estimates all converged. This excludes periods of very extreme returns. This may be the reason why the i.i.d. models and the symmetric models seem to prove adequate. In the next section the absolute violations will be considered in order to provide more insight into the performance of the different models.

8.3.2. Analysis of the absolute exceedances

The absolute exceedances, or violations, are hence investigated. Only returns below the predictions are considered, so that the absolute exceedances are defined as

$$\text{absolute exceedance} = | \text{true return} - \text{predicted VaR/ES} |,$$

$$\text{given true return} < \text{predicted VaR/ES}.$$

These are now investigated for the different models by means of histograms and summary statistics.

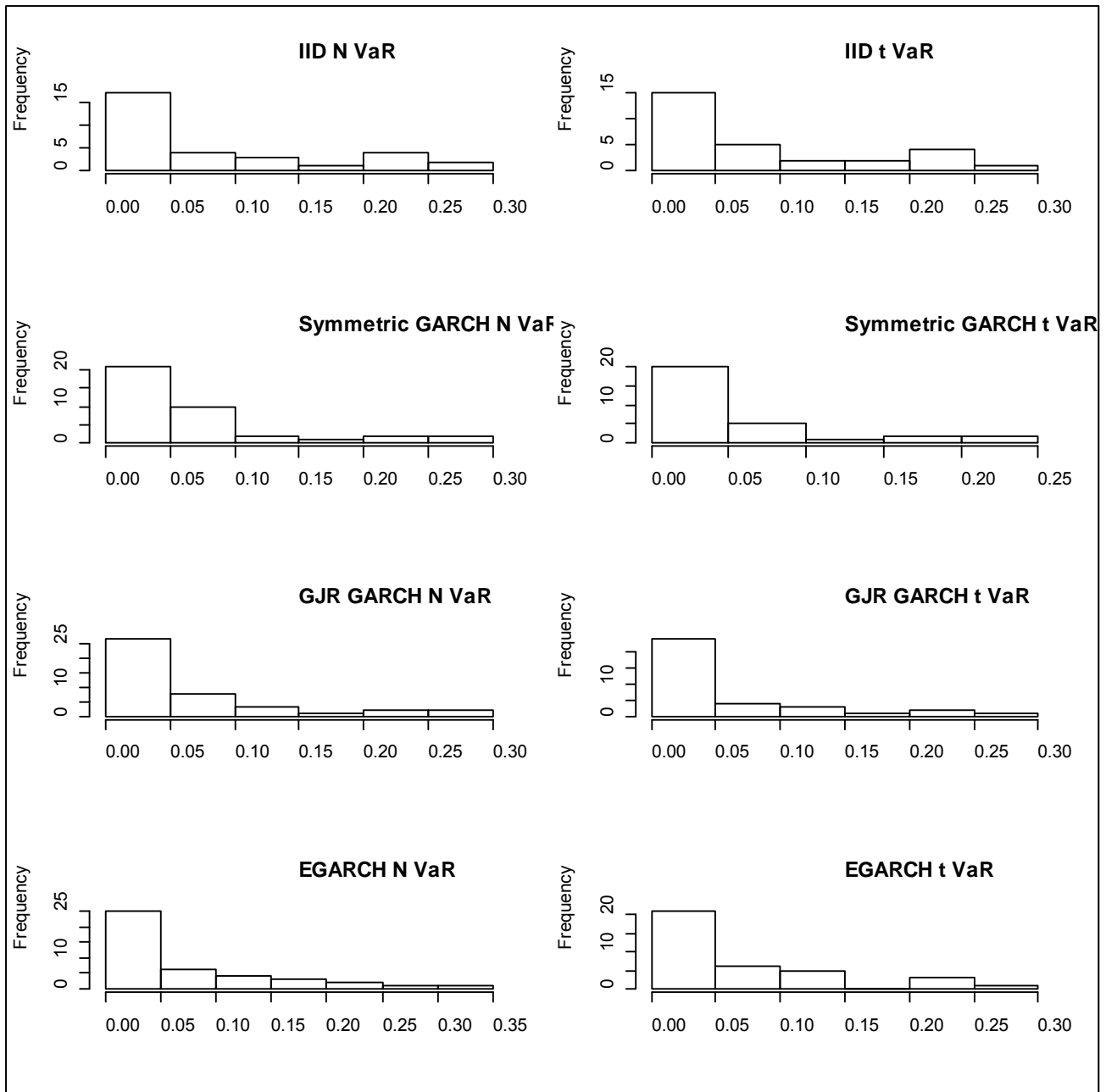


Figure 8.7: Histograms of the absolute violations from the different VaR models.

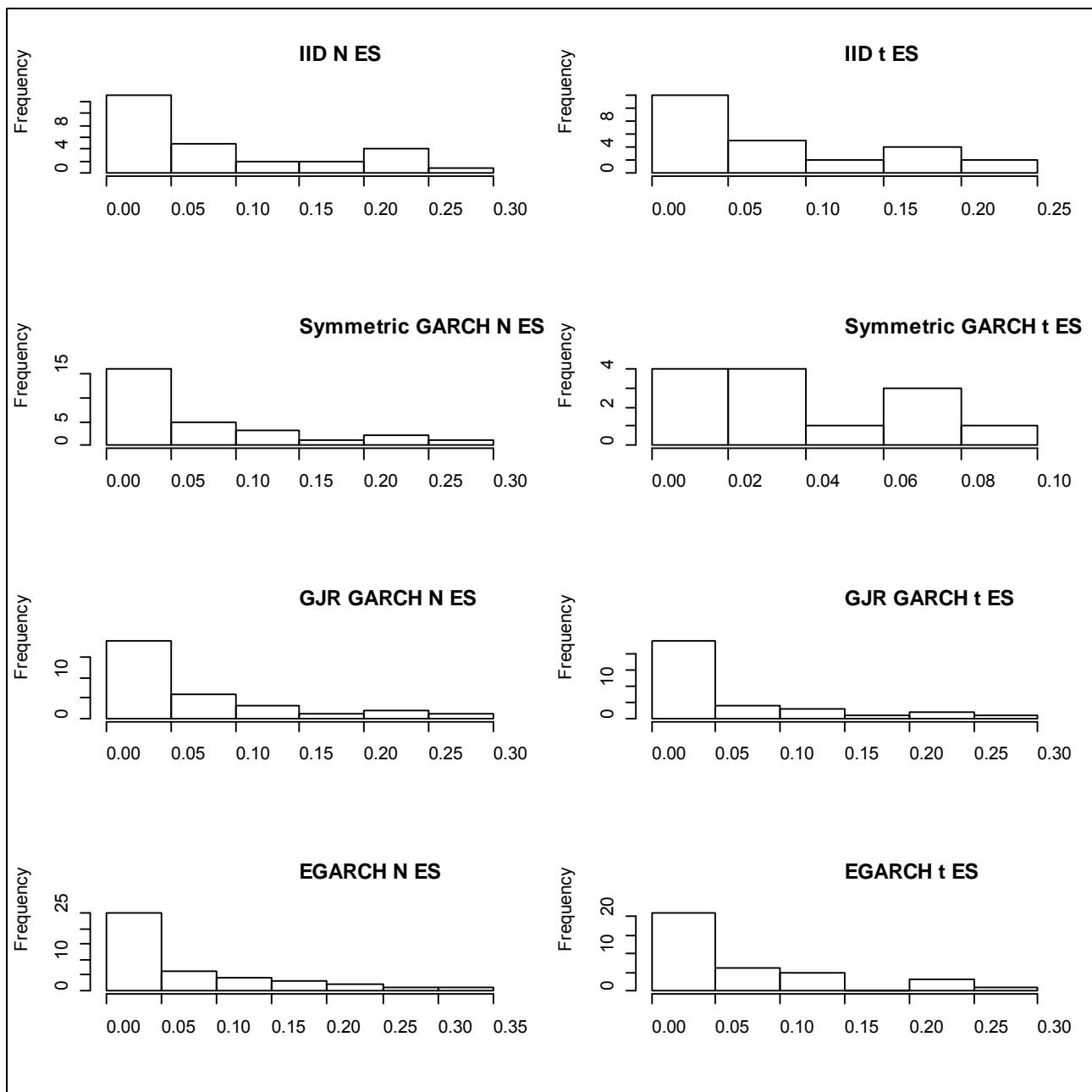


Figure 8.8: Histograms of the absolute violations from the different ES models

Table 8.3: Summary statistics for absolute violations from different VaR models.

	Mean	Std Dev	Median	Max	Min	Excess Kurtosis	Skewness
I.i.d Normal	0.085685	0.088447	0.042348	0.29207	8.18E-04	-0.3968506	0.999456
I.i.d Student t	0.085466	0.085395	0.04461	0.280858	3.48E-04	-0.5162206	0.935447
Symmetric	0.061307	0.073113	0.036264	0.276382	7.94E-05	1.7495698	1.617092

GARCH (N)							
Symmetric GARCH (t)	0.055759	0.066123	0.031808	0.23139	7.10E-04	1.104062	1.494317
GJR GARCH (N)	0.059017	0.0727	0.029454	0.294598	2.16E-03	2.7922557	1.890424
GJR GARCH (t)	0.053135	0.068028	0.028015	0.254422	1.06E-03	2.1457737	1.770752
EGARCH (N)	0.072911	0.077505	0.043278	0.30363	4.09E-03	1.5083516	1.559813
EGARCH (t)	0.067528	0.07207	0.038603	0.271357	1.24E-03	1.2535795	1.483797

Table 8.4: Summary statistics for absolute violations from different ES models.

	Mean	Std Dev	Median	Max	Min	Excess Kurtosis	Skewness
I.i.d Normal	0.083899	0.083896	0.051346	0.270794	0.000903	-0.6550071	0.864001
I.i.d Student t	0.075736	0.073995	0.054623	0.24133	0.000727	-0.6696377	0.825361
Symmetric GARCH (N)	0.066433	0.071758	0.041822	0.252821	0.00219	0.7498475	1.369085
Symmetric GARCH (t)	0.036895	0.028524	0.023862	0.084973	0.002093	-1.2723931	0.418277
GJR GARCH (N)	0.063024	0.073013	0.037592	0.273691	0.002004	1.6492685	1.630501
GJR GARCH (t)	0.053135	0.068028	0.028015	0.254422	0.001056	2.1457737	1.770752
EGARCH (N)	0.072911	0.077505	0.043278	0.30363	0.00409	1.5083516	1.559813
EGARCH (t)	0.067528	0.07207	0.038603	0.271357	0.001238	1.2535795	1.483797

The histograms in Figures 8.7 and 8.8 and tables (Tables 8.3 and 8.4) help to decide between the i.i.d. Student t model and the symmetric GARCH with student innovations, since both of these were seen to perform well previously. The histograms for the two models look similar. However, for both the VaR and the ES, the mean, median and standard deviation of the absolute violations from the symmetric GARCH with Student t innovations are lower. The maximum is lower and the minimum is higher, so the range of violations is smaller. This may provide some evidence for preferring the symmetric GARCH with Student t innovations.

Note that from these summary statistics, the GJR GARCH with Student t innovations also performs reasonably well considering the order sizes of the mean, standard deviation, median and minimum and maximum values. It seems, however, that models with t innovations are preferred over those with normal innovations.

8.4. Extreme Value Analysis

8.4.1. Exploratory analysis for first data window

In the analysis below, the first 250 data points are used for illustrative purposes and for some preliminary data exploration. That is, the *negative* of the log returns from the period 29/01/96 to 27/01/97 of the Africa Financials Index (J580) is considered (refer Figure 8.9). The reason for considering the negative of the returns is to enable application of EVT in the far right tail of the return distribution.

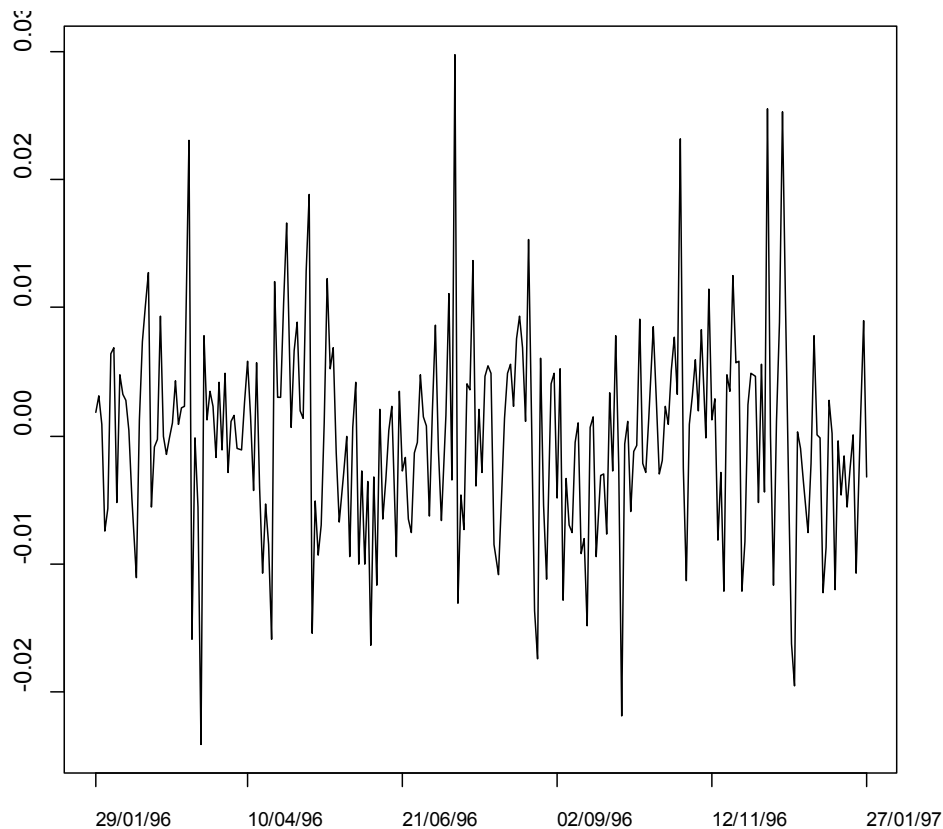


Figure 8.9: Negative of the log returns from the period 29/01/96 to 27/01/97 of the Africa Financials Index.

The histogram of the returns is shown in Figure 8.10.

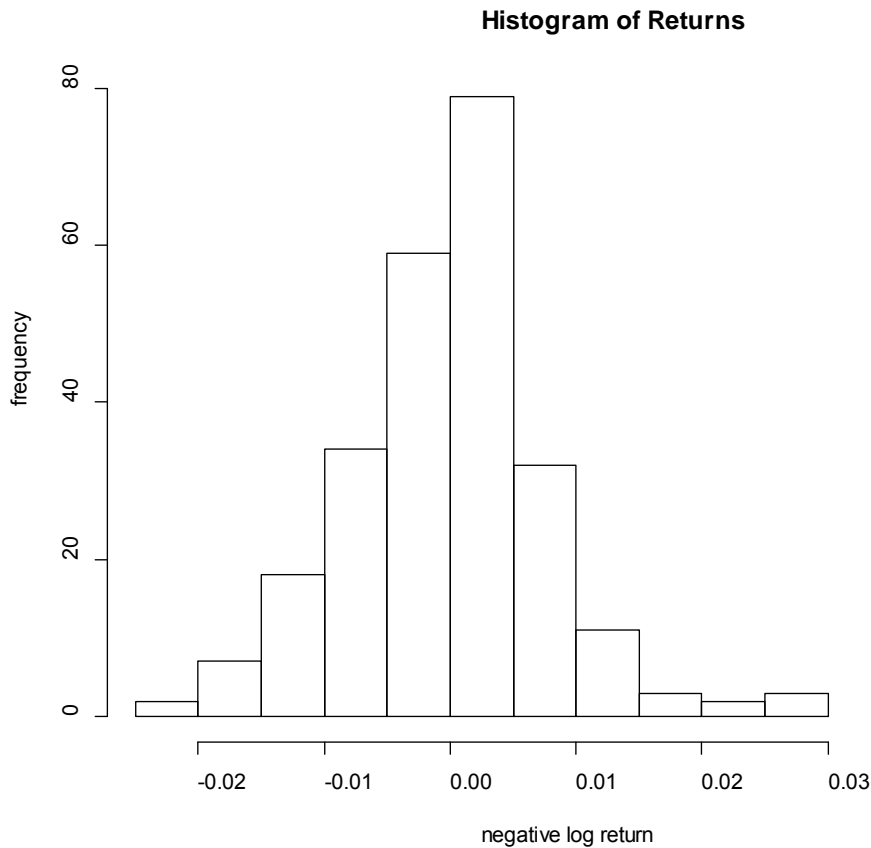
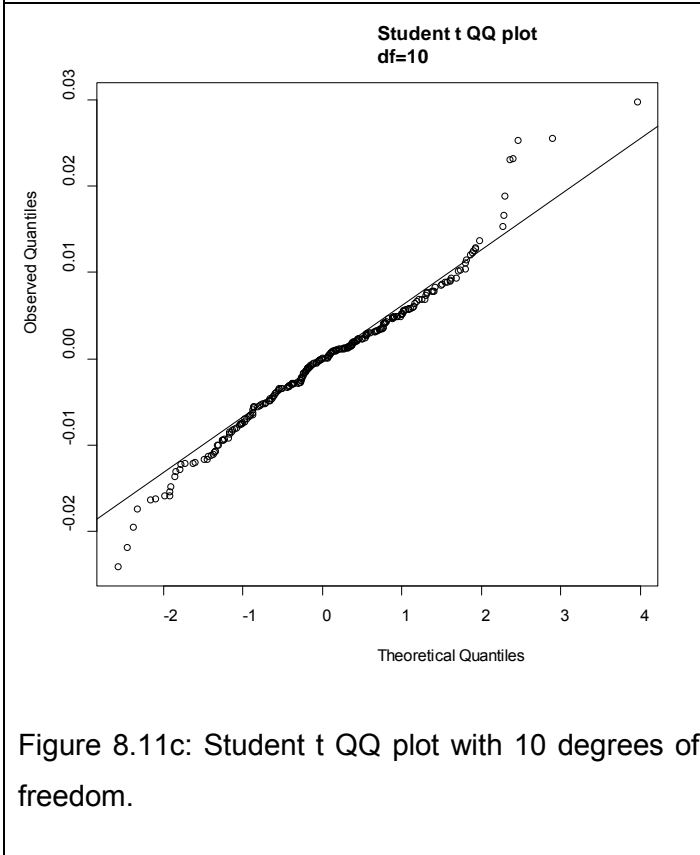
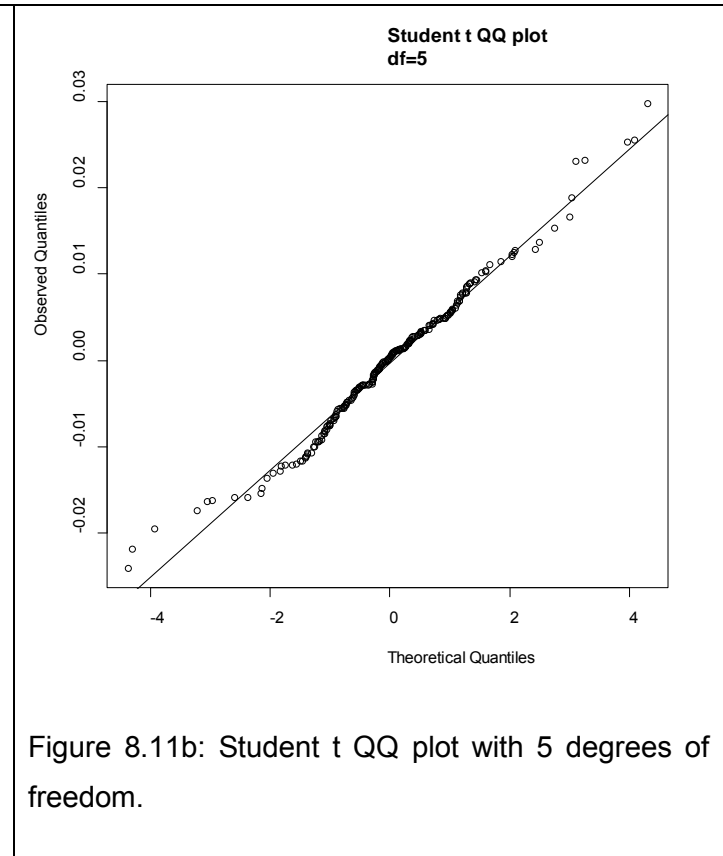
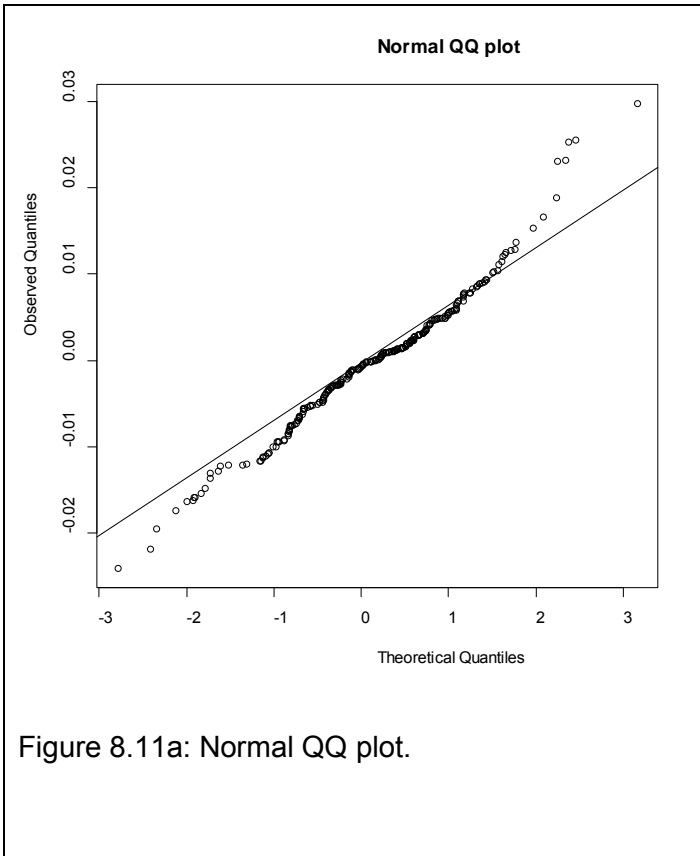


Figure 8.10: Histogram of the negative of the log returns from the period 29/01/96 to 27/01/97 of the Africa Financials Index.

The histogram is only slightly skewed to the right. A normal distribution may be appropriate, or a t distribution to allow for excess kurtosis. A Pareto distribution seems unlikely, since the tails are not so heavy. The fit of a normal distribution or Student t distribution (with 5 or 10 degrees of freedom) is assessed by the following QQ plots (Figures 8.11a-c) of the observed quantiles (the ordered returns) against the postulated, theoretical quantiles.



From Figure 8.11 the student t distribution with 5 degrees of freedom seems to give the best fit to the log returns of the three distributions considered. This is because the plot is closest to a straight line.

Since a moving data window is used, McNeil (1999) suggests taking a constant percentage of the data as the threshold value in the EVT analysis. This is merely for convenience. The suggestion is 5-10% of the sample data. The sample is very small (250 data points), so choices of 5%, 10%, 15% and 20% will be considered.

The estimates of the EVI for different exceedance percentage choices and corresponding number of exceedances (k) are summarised in Table 8.5. The table contains the moment estimator (Dekkers *et al.*, 1989; Beirlant *et al.*, 2004; refer definition 4.4.4) which is a generalisation of the Hill estimator.

The maximum likelihood (ML), method of moments (MOM) and method of probability weighted moments (PWM) based on the GPD are also provided. The confidence intervals were calculated and found to be extremely wide and uninformative. They are therefore not given. This is due to, *inter alia*, the small sample size that is considered. Figure 8.12 provides a graphical display of the values in Table 8.5.

Table 8.5: Different estimators of the EVI at various exceedance levels.

	Percentage of values considered as extreme			
	5%	10%	15%	20%
	Number of exceedances, truncated (k)			
	12	25	37	50
Moment estimator	0.0320	0.2330	0.2060	0.1950
ML estimator	-0.6557	0.0946	0.0972	0.0813
MOM estimator	-0.1412	0.0541	0.0670	0.0637
PWM estimator	-0.2878	0.0762	0.0655	0.0536

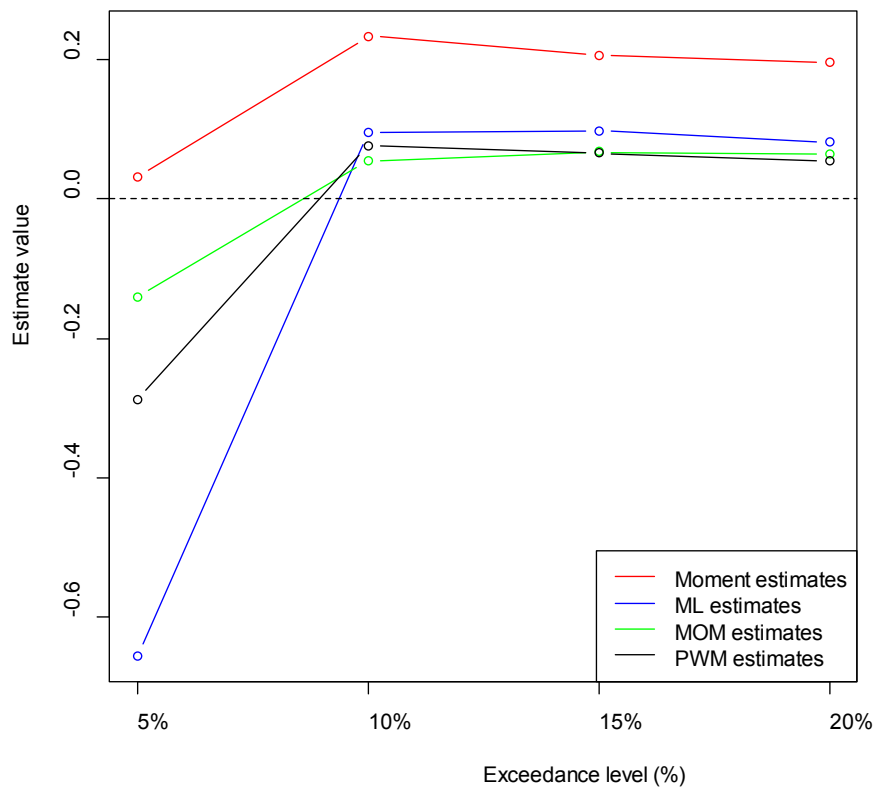


Figure 8.12: Plot of different estimators of the EVI at various exceedance levels

The estimators at 5% exceedance level are very variable. The ML, MOM and PWM estimators are negative, indicating a lighter tail than exponential behaviour of the tail index. It is suggested not to consider 5% as an option for the threshold, since this results in only about 12 exceedances which are too few observations for EVT to be applied sensibly. The values of the EVI seem to stabilize slightly for the different estimation procedures for the 10%, 15% and 20% levels. The moment estimators are markedly higher than the other estimators. The latter tend to be close to zero, indicating membership to the Gumbel class of distributions. The important point to remember is that these estimators are based on asymptotic results, whereas the sample sizes here are very small. All estimators should therefore be interpreted with care.

The suggestion is to consider 15% and 20% as possible threshold levels, since this will lead to more observations in the tail to which EVT can be applied. Furthermore, in the rest of the study the Moment estimators and ML estimators will be considered for the purposes of VaR and ES calculations. The choices are somewhat subjective. ML estimators are generally used in the literature and the Moment estimator provides a useful variation in that it is a consistent estimator for all values of the EVI.

Since the estimators in Table 8.5 are positive for exceedances equal to 25 or larger, it is also informative to examine the corresponding Hill estimators. These are summarised in Table 8.6.

Table 8.6: Hill estimators of the EVI at the relevant exceedance levels.

	Percentage of values considered as extreme		
	10%	15%	20%
	Number of exceedances, truncated (k)		
	25	37	50
Hill estimator	0.4418	0.5150	0.6177

The Hill estimators increase as the number of exceedances increases. The Hill estimators are higher than the estimators in Table 8.5 and should be interpreted with caution since they are only valid under a positive EVI assumption.

The mean excess plot is provided in Figure 8.13, where u is the threshold return multiplied by 100. The plot is constant between 0 and 0.8, so the threshold return can possibly be somewhere between 0 and 0.008 (0.8%). For 15% exceedances, the threshold return is 0.006698 and for 20% exceedances, the threshold return is 0.005089. Therefore the mean excess plot provides some justification for these threshold choices.

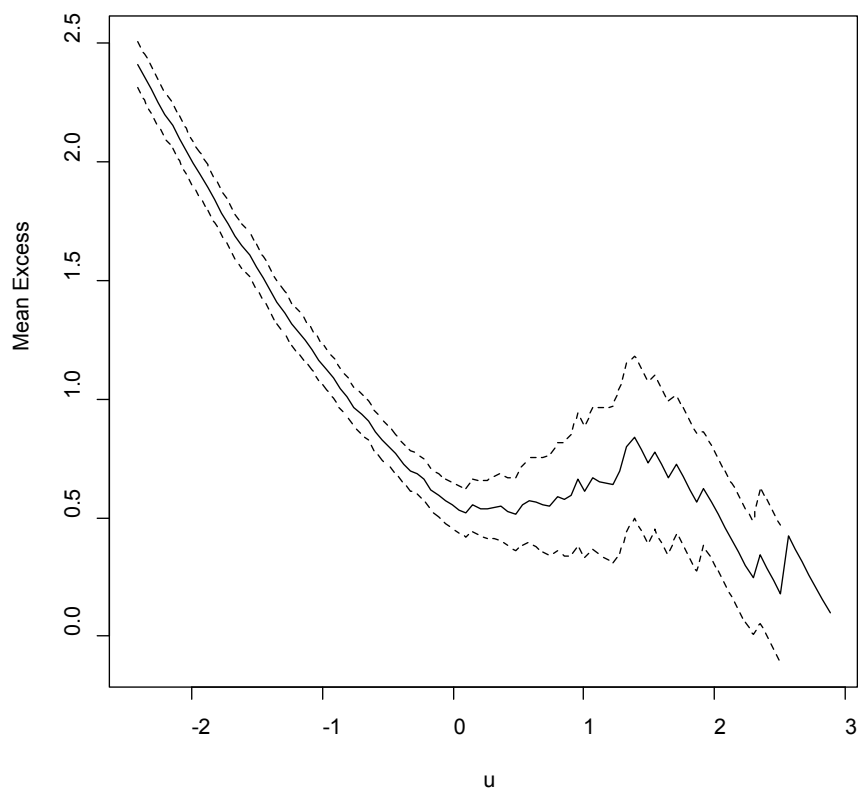


Figure 8.13: Mean excess plot for different threshold return levels.

QQ plots are used to assess the reasonability of a GPD for the exceedances. The reasoning is as follows.

Let

$$u = G(y) = 1 - \left(1 + \frac{\gamma y}{\sigma}\right)^{-\frac{1}{\gamma}}$$

and solve for y to obtain

$$y = \frac{\sigma}{\gamma} [(1-u)^{-\gamma} - 1].$$

The constant $\frac{\sigma}{\gamma}$ can be treated as a scale parameter, therefore the appropriate QQ plot of the theoretical against the observed quantiles has the form

$$\left\{ \left(1 - \frac{i}{n+1}\right)^{-\hat{\gamma}} - 1; y_{in} \right\}, i = 1, \dots, n$$

where y_{in} denotes the ordered exceedances ($n=12,25,37$ or 50 as appropriate) and $\hat{\gamma}$ is estimated from the data (either the moment, ML, MOM or PWM estimator in Table 8.5).

The QQ plots are given in Figures 8.14 to 8.17. The plots are similar for the different estimators. Note that at 5% exceedance level, the slope of the QQ plots for the ML, MOM and PWM estimators are negative, because the estimate of the EVI (Table 8.5) is negative. The QQ plots are linear at the start, providing evidence for the appropriateness of the GPD assumption for the exceedances. The points deviate from linearity toward the right of the plots. However, this is to be expected because of the scarcity of the data in these areas. Not too much reliance should be placed on the behaviour of the tails of the QQ plots.

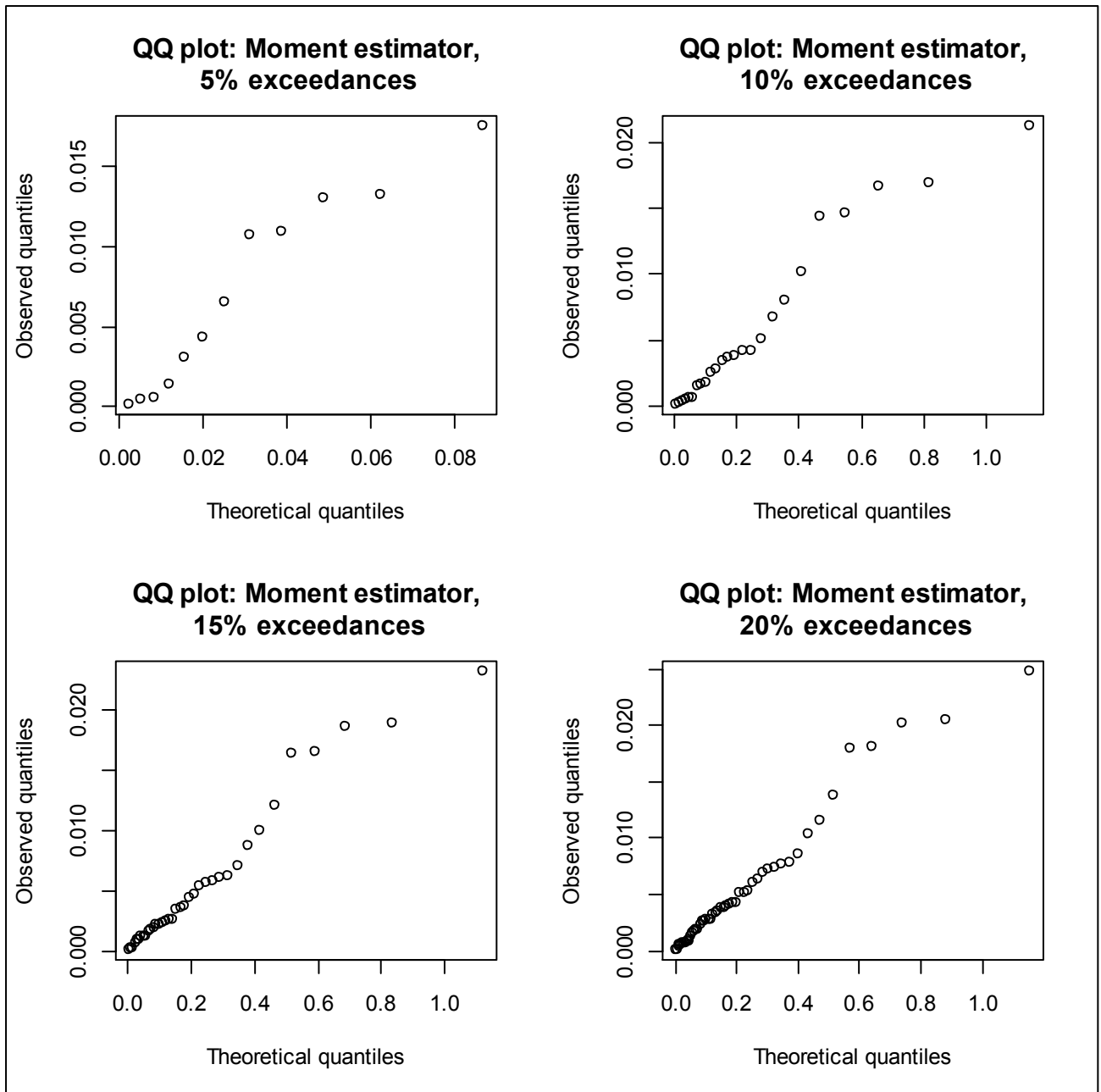


Figure 8.14: QQ plots for the Moment estimator for different exceedance levels.

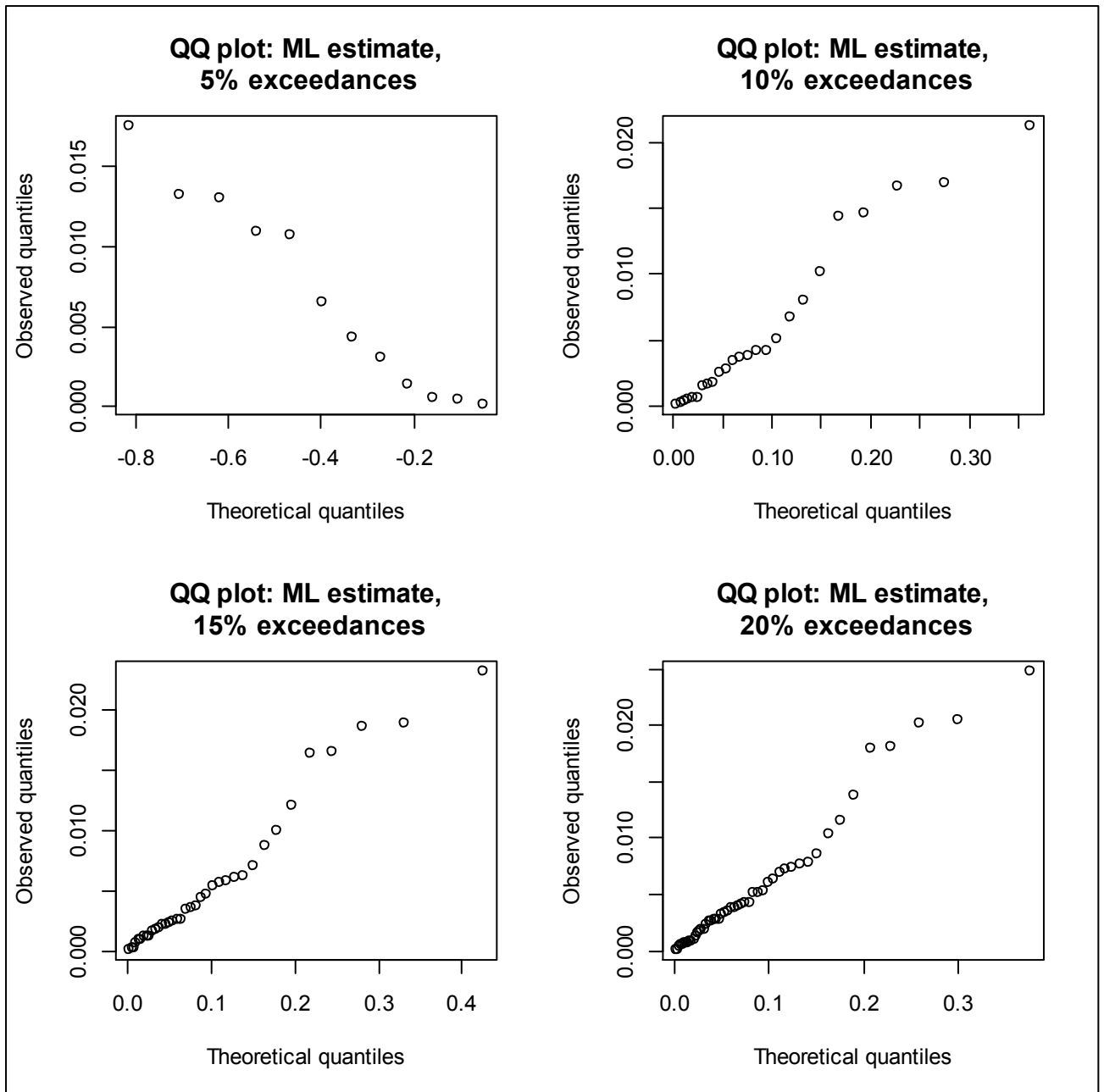


Figure 8.15: QQ plots for the ML estimator for different exceedance levels.

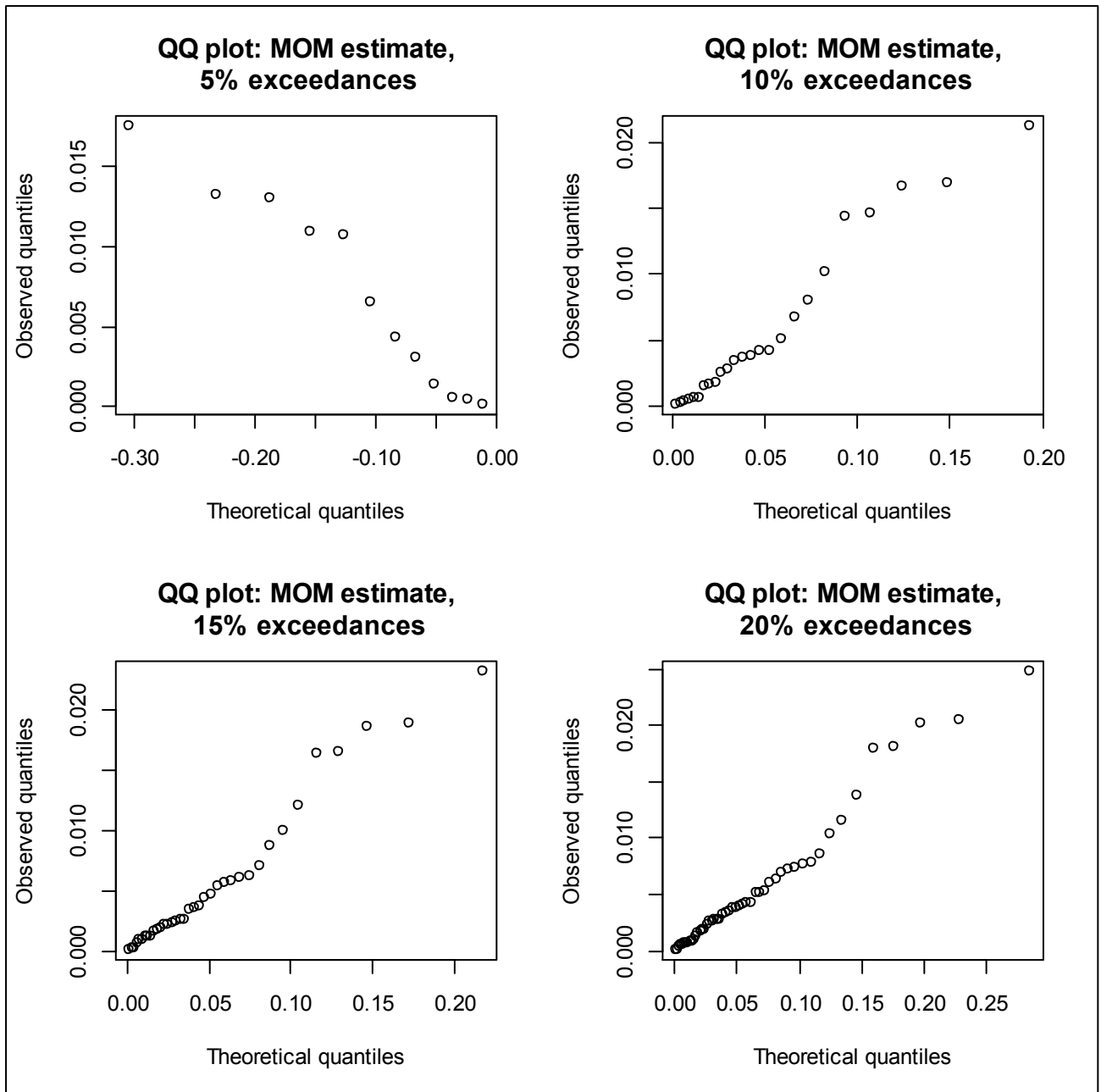


Figure 8.16: QQ plots for the MOM estimator for different exceedance levels.

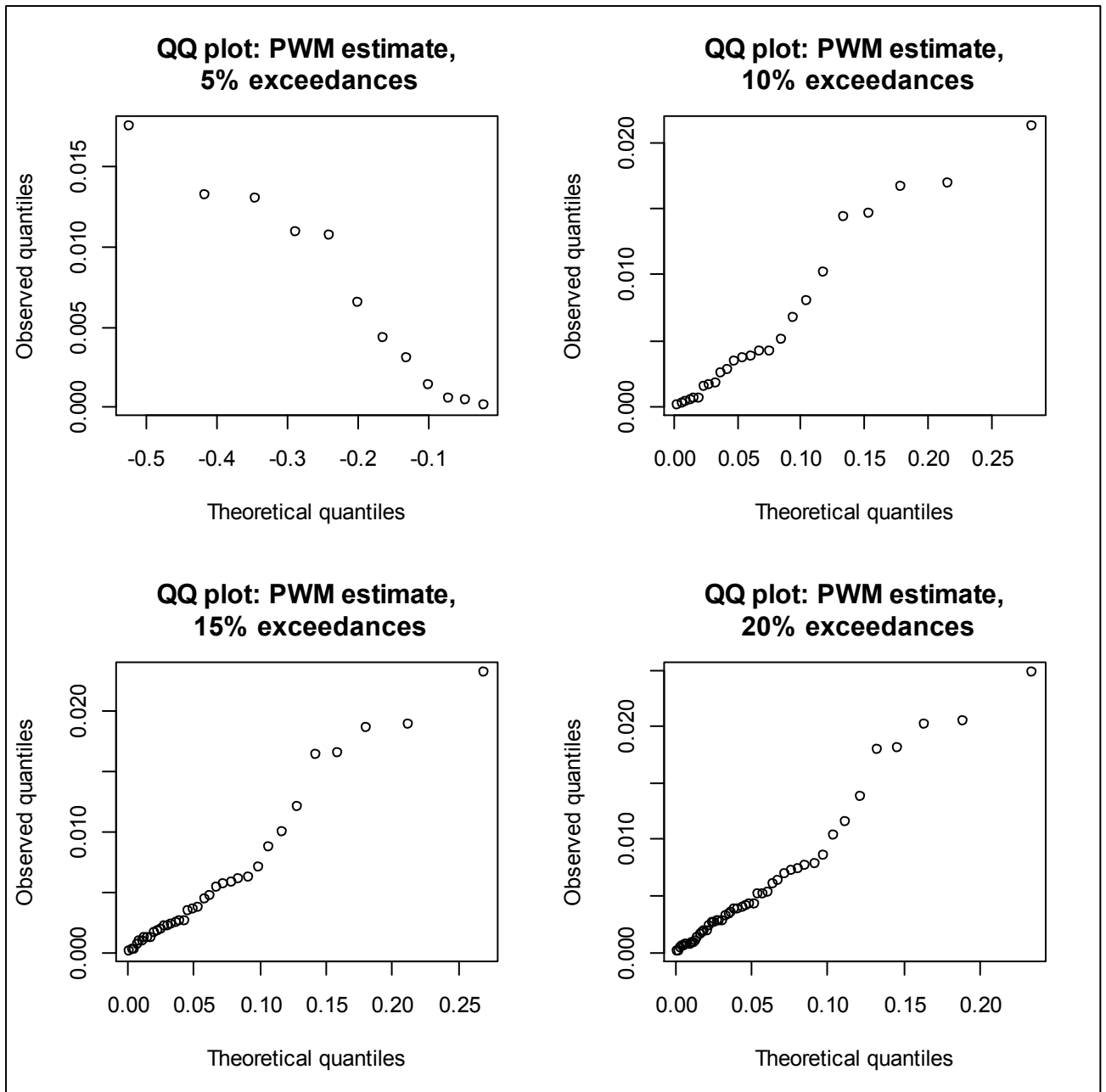


Figure 8.17: QQ plots for the PWM estimator for different exceedance levels.

The insights gained from this exploratory analysis of the first data window will next be used to estimate VaR and ES over the whole data series.

8.4.2 VaR and ES over the whole data series

Extreme value theory is applied to a moving data window of length 250 days. The parameters are therefore re-estimated each day based on the previous 250 days of data. The data is the negative of the log returns of the Africa Financials Index (J580) from 29/01/1996 to 30/04/2013. A threshold of 20% is used throughout. The ML estimation method is used for the estimation of the parameters of the GPD. The estimator of the EVI from this method (the gamma parameter in the GPD) is referred to as the ML estimator. As an additional estimator of the EVI, the Moment estimator (definition 4.4.4) is also used. These two estimators of the EVI over the sample space, using a moving 250 days data window, are shown in Figure 8.18. The ML estimator of the sigma parameter for the GPD is plotted in Figure 8.19. This is later used in the calculation of the VaR and ES estimates.

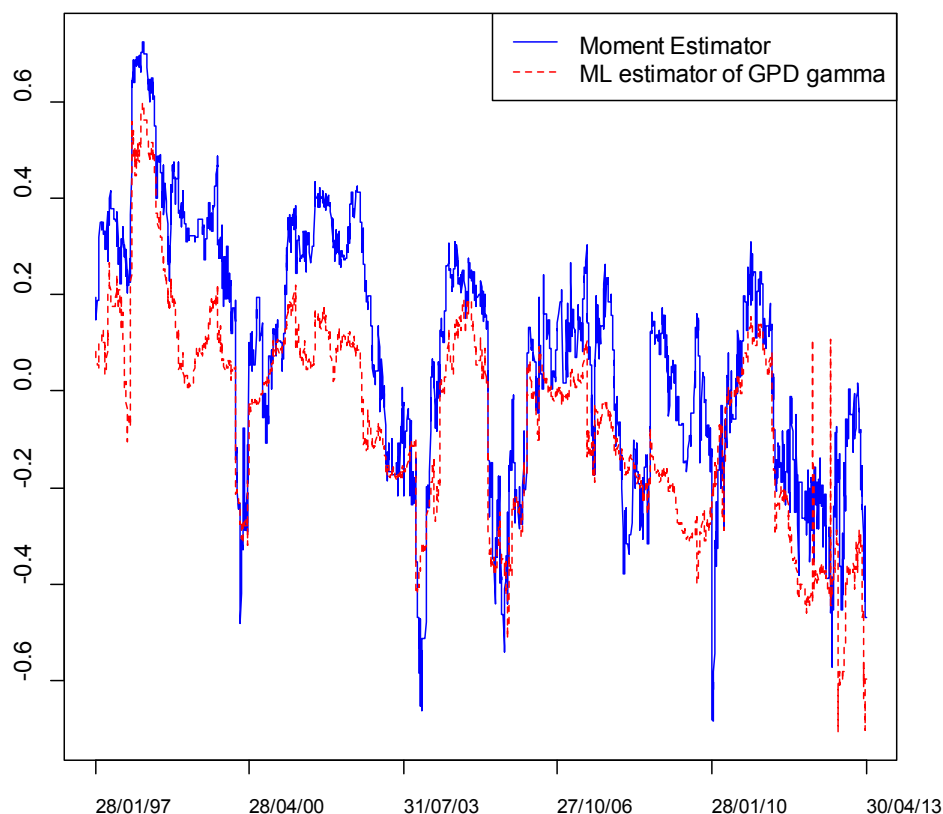


Figure 8.18: Moment estimators and ML estimators (for the GPD) of the EVI based on a moving data window of 250 days.

It can be seen in Figure 8.18 that the Moment estimators are mostly larger than the ML estimators, except at large negative values where the order is sometimes reversed.

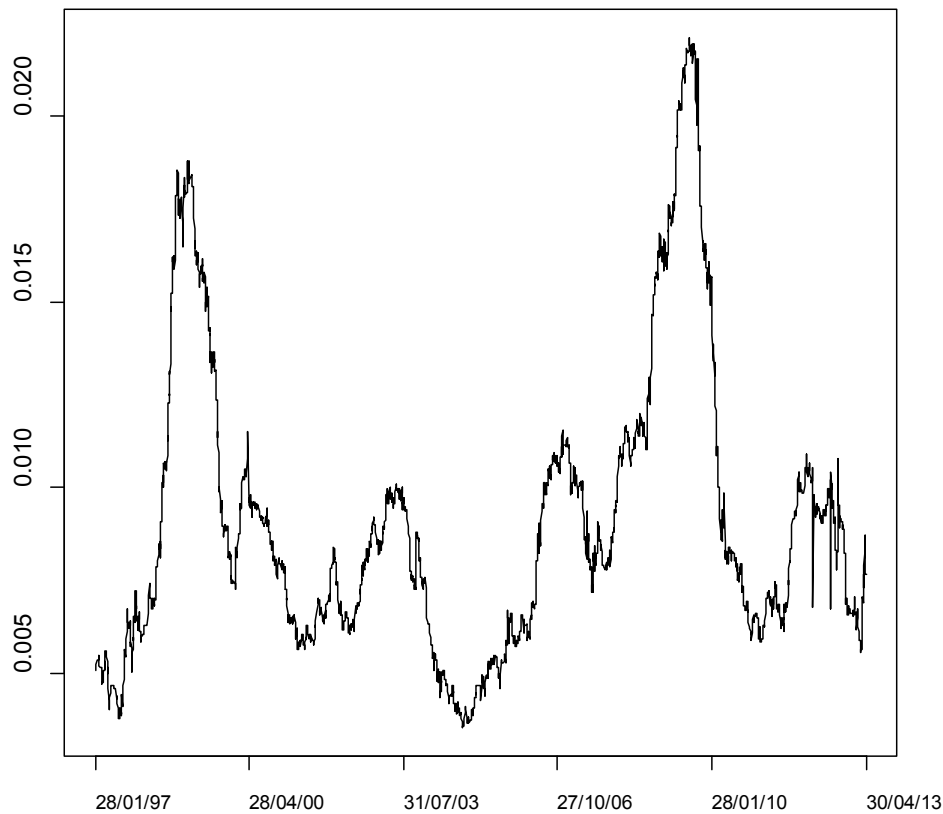


Figure 8.19: ML estimators of the GPD parameter sigma, based on a moving data window of 250 days.

The ML estimators of the GPD parameter sigma are very variable over time (Figure 8.19). The ratios of the ML estimator for sigma to the ML estimator for the EVI are plotted in Figure 8.20. These are still extremely variable.

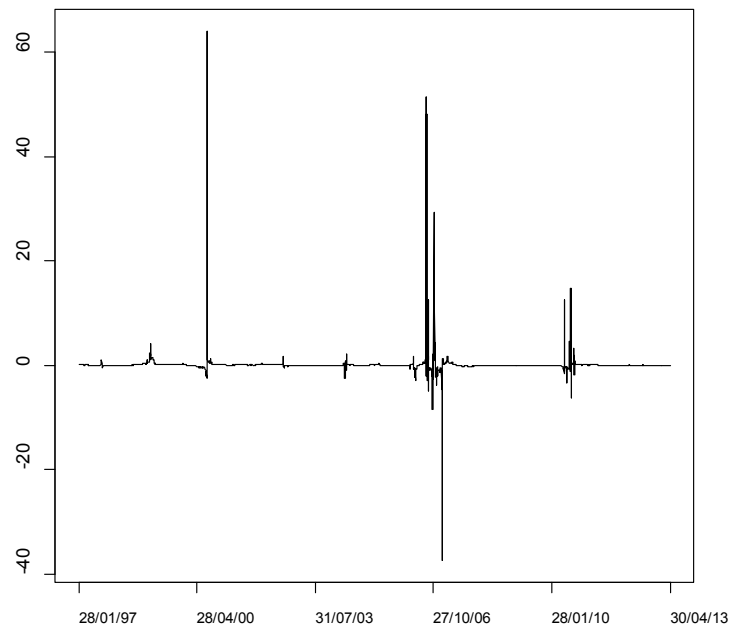


Figure 8.20: Ratios ratios of the ML estimator for sigma to the ML estimator for the EVI.

Hence, the VaR and ES estimates are calculated. First it is done for one day horizons and then extended to ten day forecasts.

8.4.2.1. One day VaR and ES estimates

The one day 99% VaR and ES estimates are based on Results 4.6.1 and 4.6.2 respectively. They are plotted together with the one day log returns in Figure 8.21. Note that the original returns are plotted (not the negative of the returns). Instead, the VaR and ES estimates are multiplied by (-1) to abide by convention. There are a total of 4313 log return values. Therefore, starting on day 251, there are $4313 - 251 + 1 = 4063$ true one day returns against which the VaR and ES estimates can be compared. The expected number of 99% VaR and ES violations is then equal to 40.63.

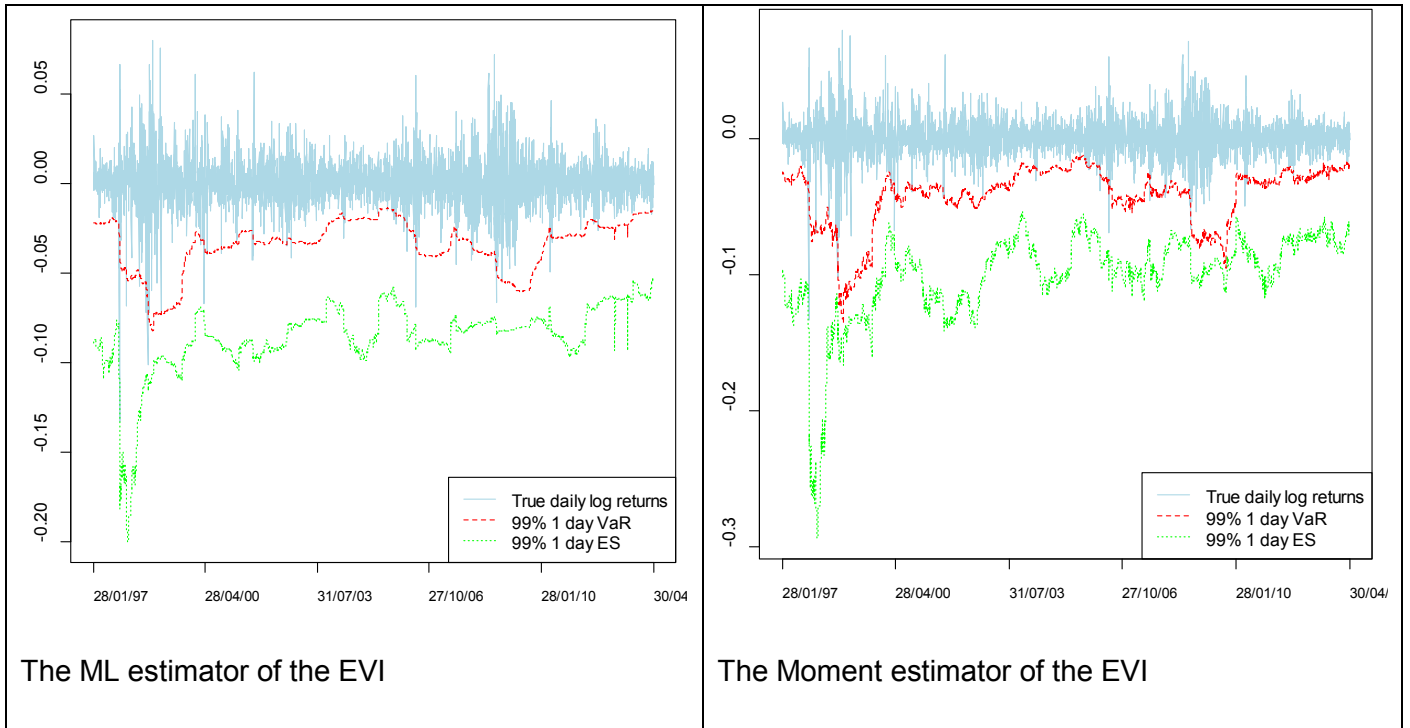


Figure 8.21: One day POT VaR and ES forecasts, using different ML estimators of the EVI.

The violations are summarised in Table 8.7 and Figure 8.22.

Table 8.7: One day VaR and ES violations for the POT estimate, using the ML estimate and the Moment estimate respectively (over the entire sample period).

	VaR forecast	ES forecast	Expected
POT (ML estimate)	53	1	40.63
POT (Moment estimate)	41	0	40.63

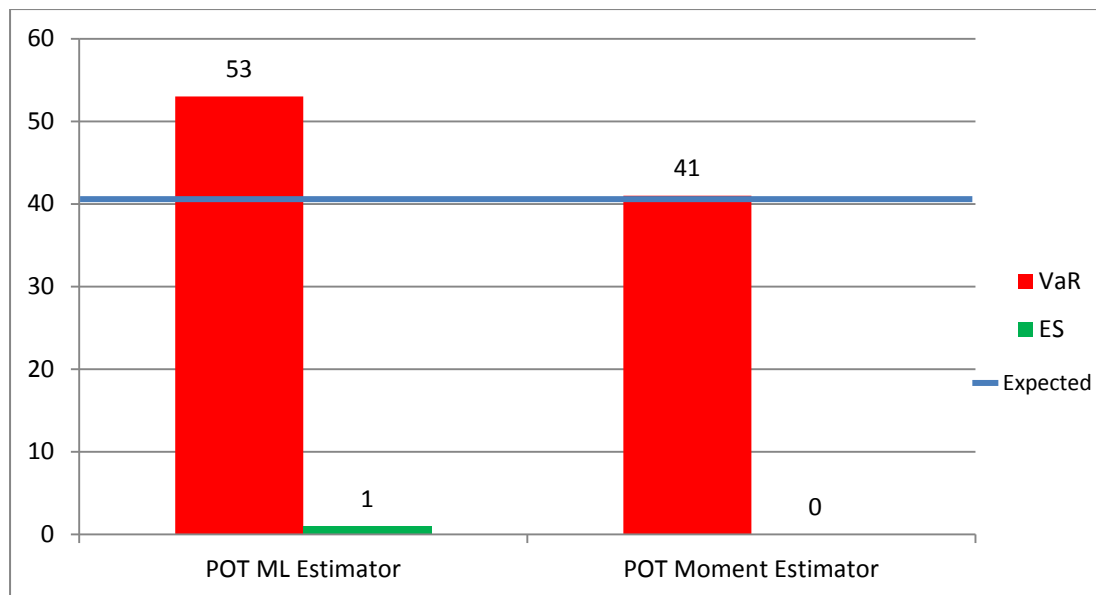


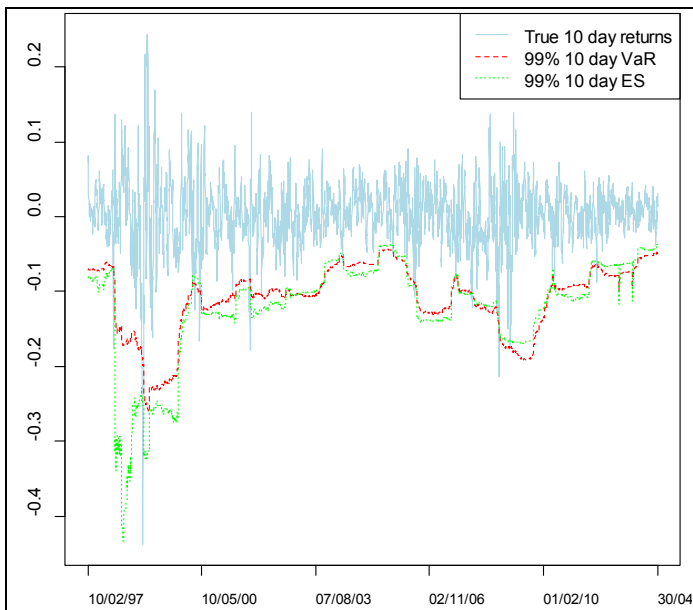
Figure 8.22: Bar chart of one day VaR and ES violations (observed and expected) for the POT estimate, using the ML estimate and the Moment estimate respectively (over the entire sample period).

The moment estimator yields fewer violations than the ML estimator. It is interesting that the ES estimates are in fact overly conservative and probably not of much use in this situation.

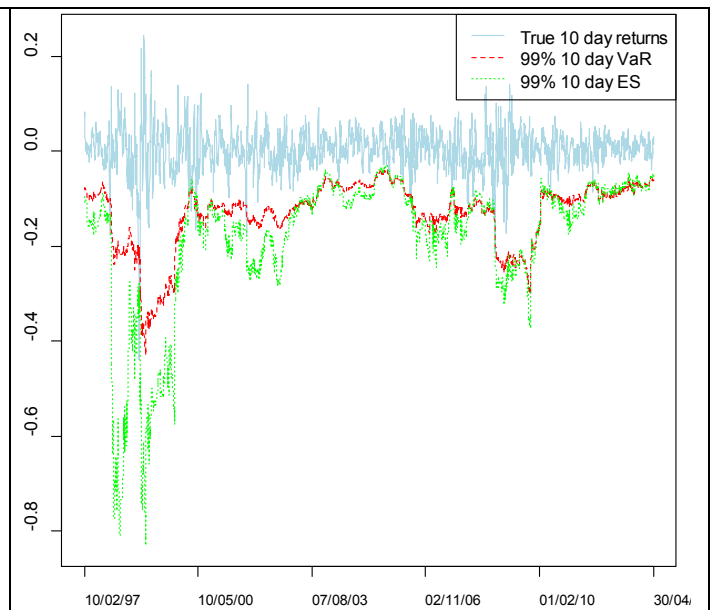
The next section addresses the important issue of scaling these one day estimates to ten day estimates.

8.4.2.2. Ten day VaR and ES estimates

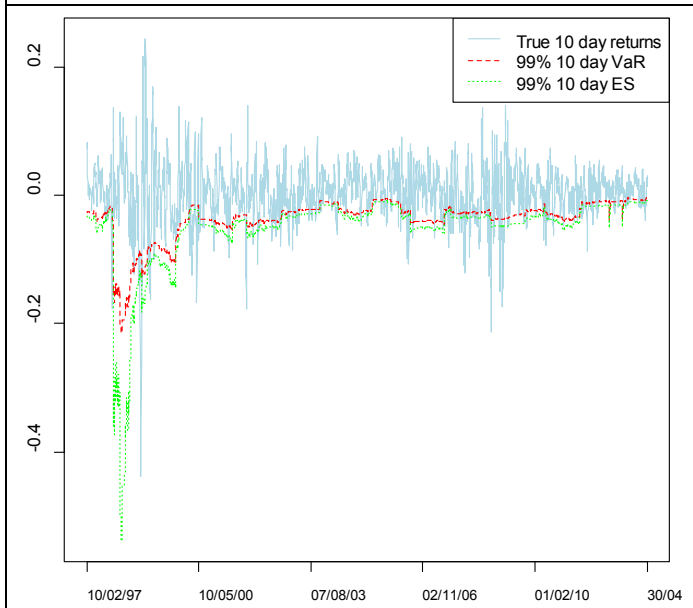
The three possibilities of scaling VaR and ES from one day estimates to ten day estimates (refer Section 4.8 of Chapter 4) are empirically investigated in this section. These are the square root of time rule, the alpha root rule and the empirically established scaling laws. The estimates over the entire sample period are provided in Figure 8.23. Figure 8.24 and 8.25 pertain to the empirically established scaling laws. The results for all the models are summarised in Table 8.7 and Figure 8.26.



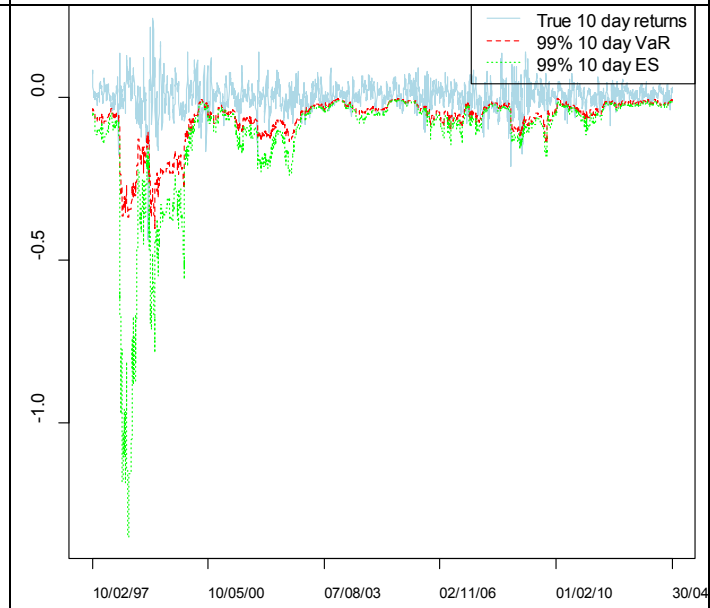
The ML estimate of the EVI and the square root of time scaling rule.



The Moment estimate of the EVI and the square root of time scaling rule.

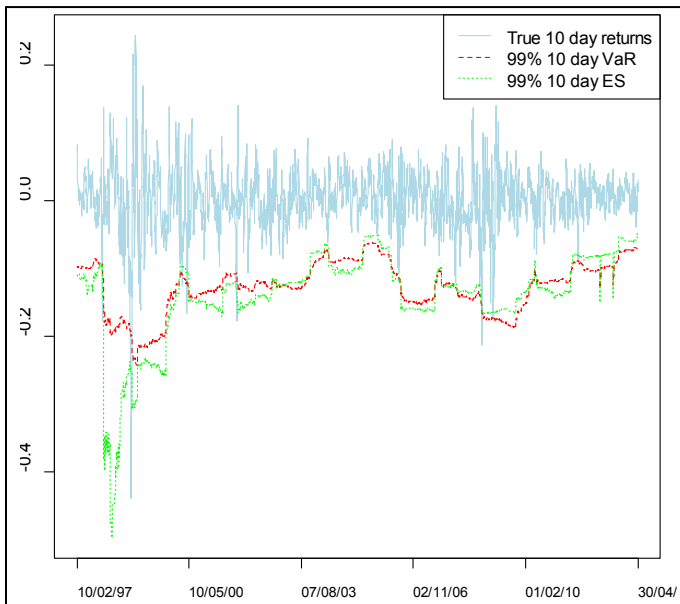


The ML estimate of the EVI and the alpha root scaling rule.

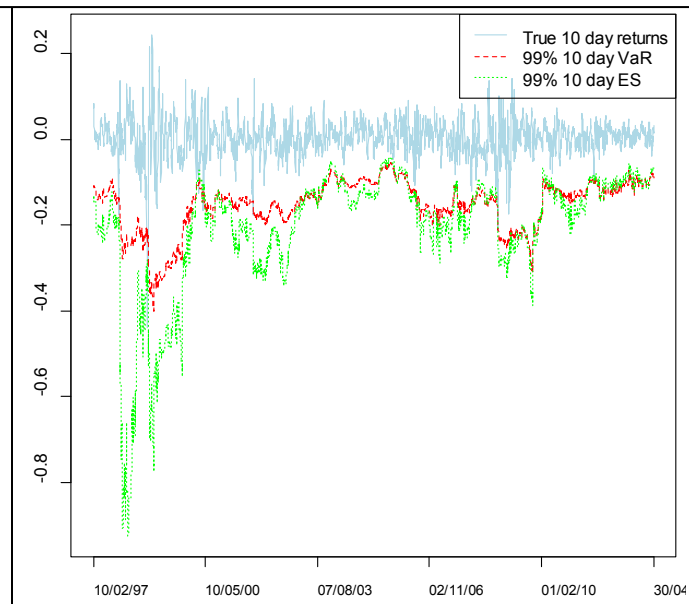


The Moment estimate of the EVI and the alpha root scaling rule.

Figure 8.23: Ten day POT approach VaR and ES forecasts using different estimates of the EVI and different scaling methods.



The ML estimate of the EVI and the empirical scaling rule (scaling factors of Figure 8.25 were used).



The Moment estimate of the EVI and the empirical scaling rule (scaling factors of Figure 8.25 were used).

Figure 8.23 (continued): Ten day POT approach VaR and ES forecasts using different estimates of the EVI and different scaling methods.

The procedure for the empirically established scaling laws (refer section 4.8.3) was implemented as follows. The standard deviation based on the past 250 days' returns was calculated and compared to the percentiles of the historical distribution. The function used for the linear interpolation is plotted in Figure 8.24. Figure 8.25 gives the scaling factors used for the VaR and ES estimates in Figure 8.23.

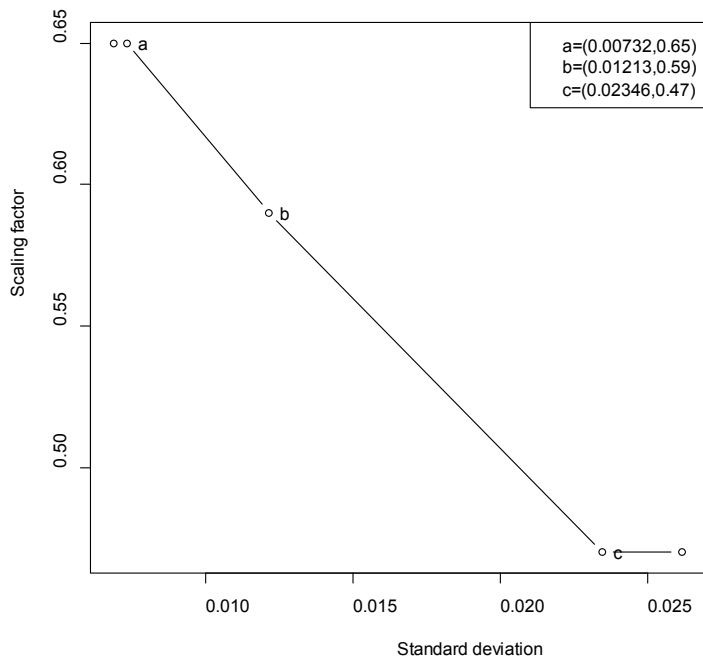


Figure 8.24: Function used for linear interpolation in the determination of the scaling factors to convert 1 day estimates to 10 day estimates.

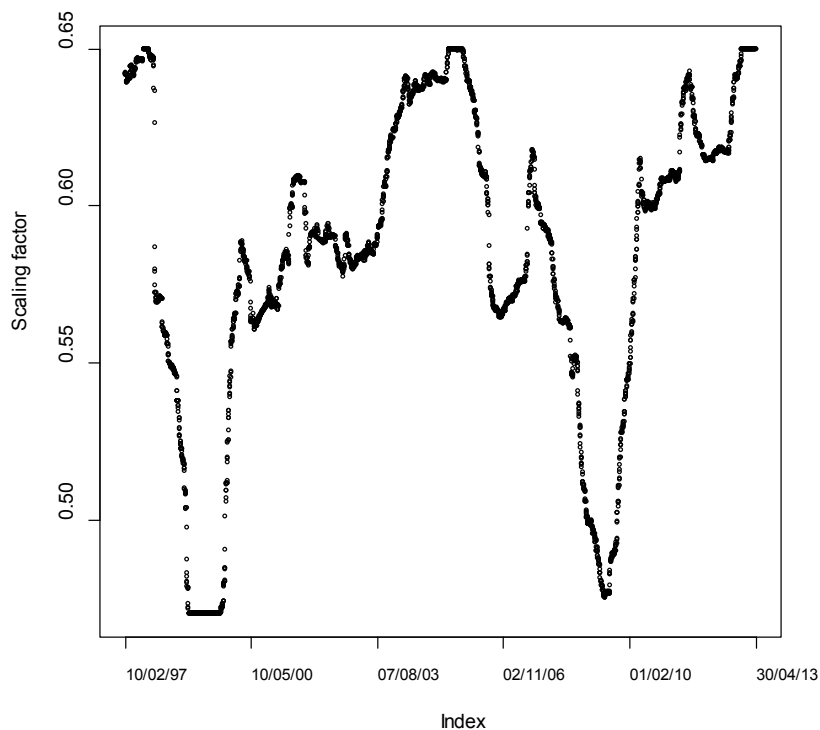


Figure 8.25: Scaling factors used on different days for the VaR and ES estimates.

The violations from all the methods are summarised in Table 8.7 and Figure 8.26.

Table 8.7: Actual violations from the ten day POT approach VaR and ES forecasts, using both the ML and Moment estimates of the EVI and different scaling methods.

	VaR forecast	ES forecast	Expected
POT (ML, Square root)	42	49	40.45
POT (Moment, Square root)	29	28	40.45
POT (ML, Alpha root)	660	487	40.45
POT (Moment, Alpha root)	423	312	40.45
POT (ML, Empirical)	22	23	40.45
POT (Moment, Empirical)	15	12	40.45

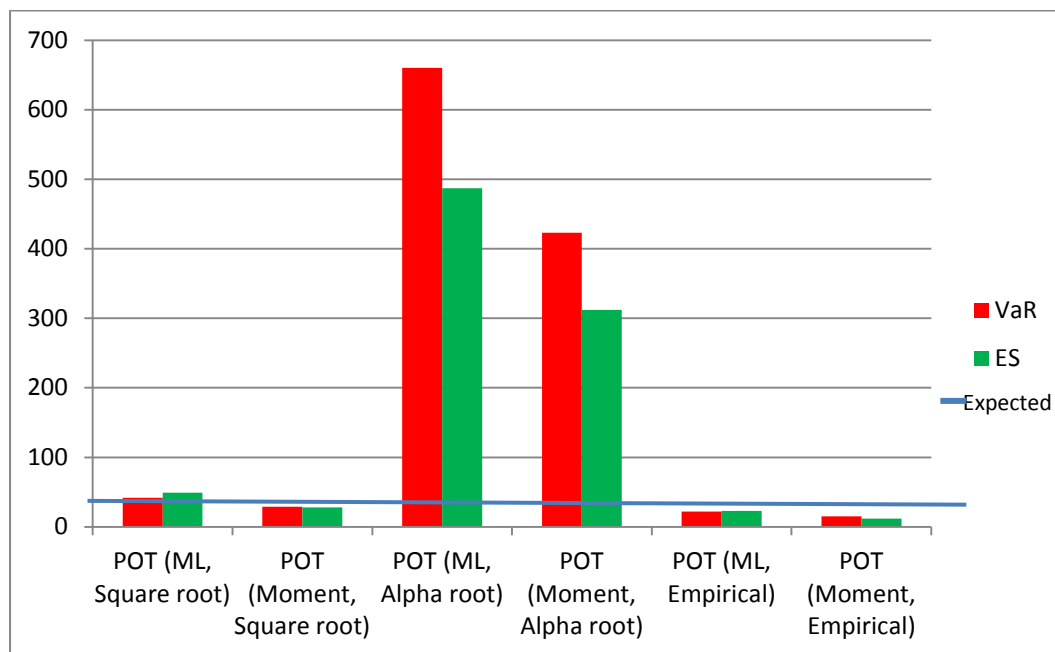


Figure 8.26: Bar Chart of actual violations from the ten day POT approach VaR and ES forecasts, using both the ML and Moment estimates of the EVI and different scaling methods.

The alpha root rule yields by far the most violations and performs the poorest. The empirical scaling rule yields the lowest number of violations. The square root of time rule seems to produce slightly more violations than the empirical scaling rule. The Moment estimator

results in fewer violations than the ML counterparts. It is interesting that the ES forecasts are in this case not consistently lower than the VaR forecasts.

In order to further compare the relative performance of the different scaling methods, Figure 8.27 plots the VaR and ES forecasts from the different scaling laws for the ML and the Moment estimates respectively.

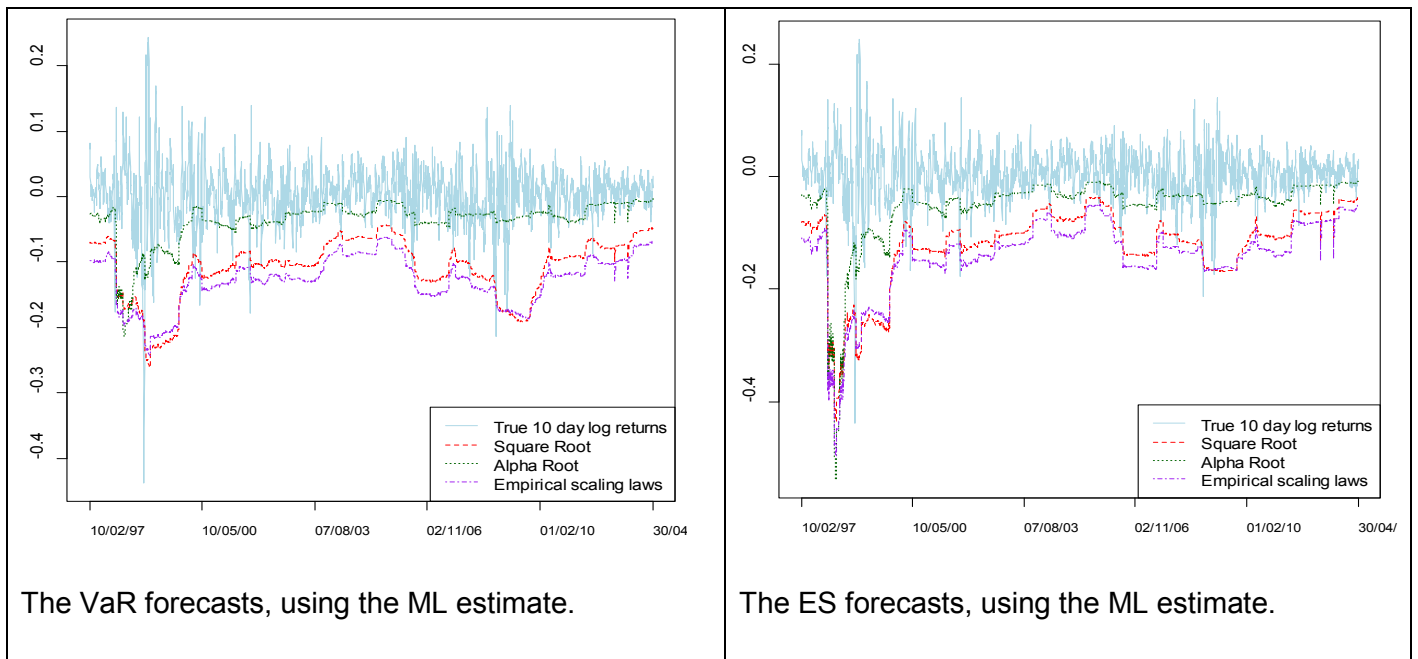


Figure 8.27: Relative comparison of how the forecasts from the different scaling rules differ with respect to the estimate of the EVI used (for VaR and ES respectively).

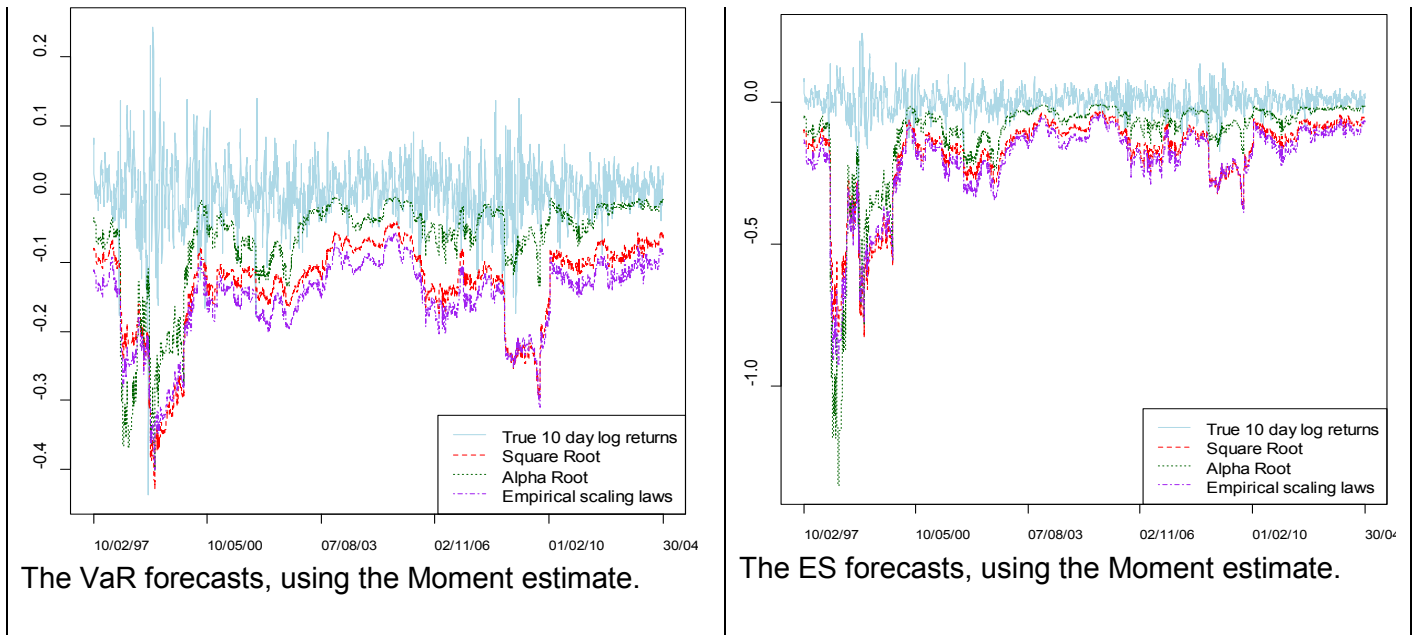


Figure 8.27 (continued): Relative comparison of how the forecasts from the different scaling rules differ with respect to the estimate of the EVI used (for VaR and ES respectively).

In Figures 8.27 the empirical scaling law produces the lowest forecasts (the most conservative, leading to the least number of violations). The square root rule yields the second lowest forecasts, with the alpha root rule yielding substantially higher forecasts.

8.4.3. Augmenting GARCH forecasts with EVT forecasts

The GARCH models do not converge over the entire sample period, as was seen in Section 8.3. This is mainly due to the presence of extreme observations in the estimation periods. (Refer Chapter 6 for the detailed discussion hereof.) This leads to gaps in the GARCH forecasting models, where convergence is not possible.

These gaps in the GARCH forecasts can be filled with EVT estimates, since the latter converge over the entire sample period. The EVT models are especially designed to work where the GARCH models fail. The EVT models are by design well suited to samples containing extreme observations. This is precisely what is needed during periods where the GARCH models do not converge. It is therefore sensible to augment the GARCH forecasts with EVT forecasts to provide forecasts over the entire sample period and to compare these results with those of earlier models.

The empirically established scaling laws (refer Section 4.8.3) will be used to convert the one day POT VaR and ES estimates to ten day estimates. The ML estimate and the moment estimate of the EVI will be considered.

The results are provided in Figures 8.28 to 8.32. Figures 8.28 and 8.29 show the enlarged versions for the two best performing models, based on the number of VaR violations (refer Table 8.8). The results for the other models are given on reduced scale. The results for the symmetric GARCH models are given in Figure 8.30 and those for GJR- and E-GARCH models in Figures 8.31 and 8.32 respectively.

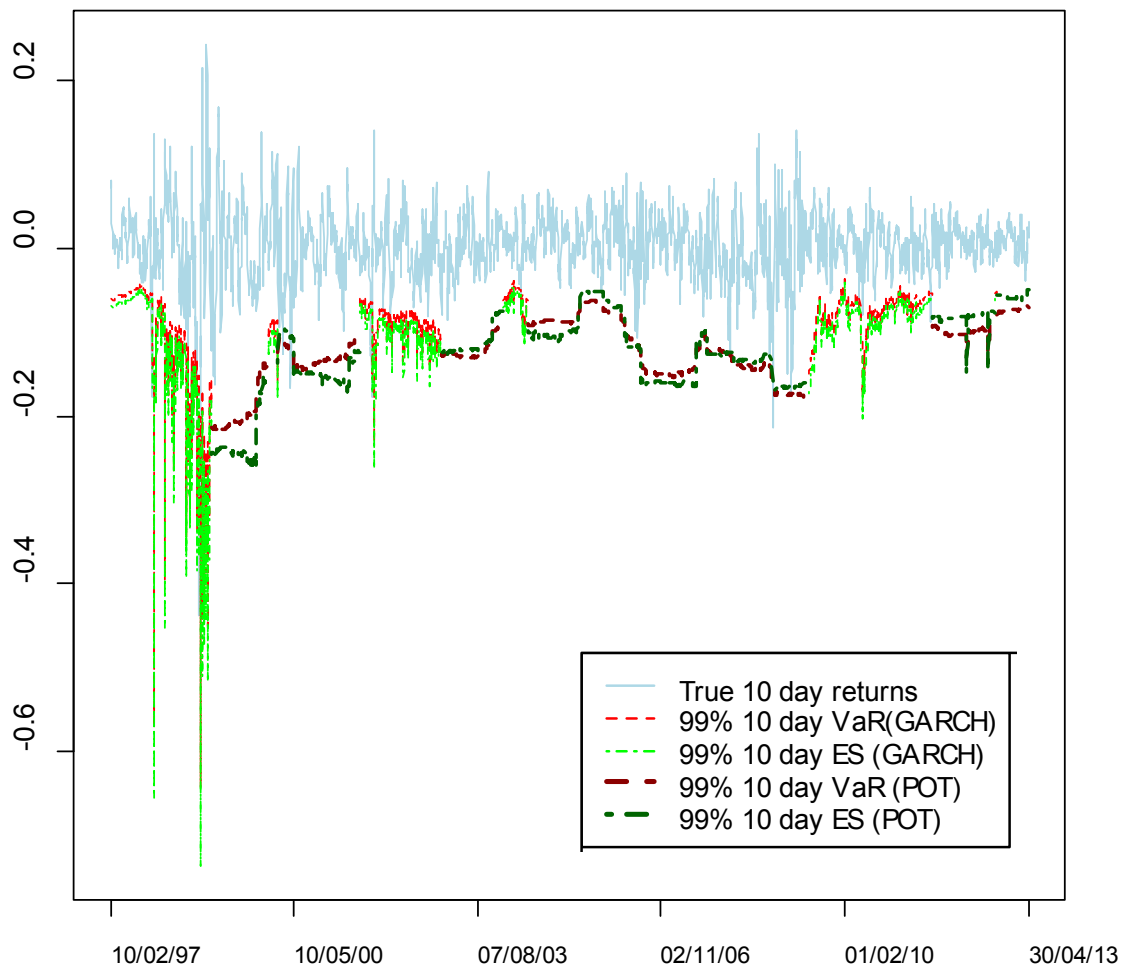


Figure 8.28: Forecasts from the Symmetric GARCH model (normal innovations) augmented with the POT EVT model (ML estimate of the EVI).

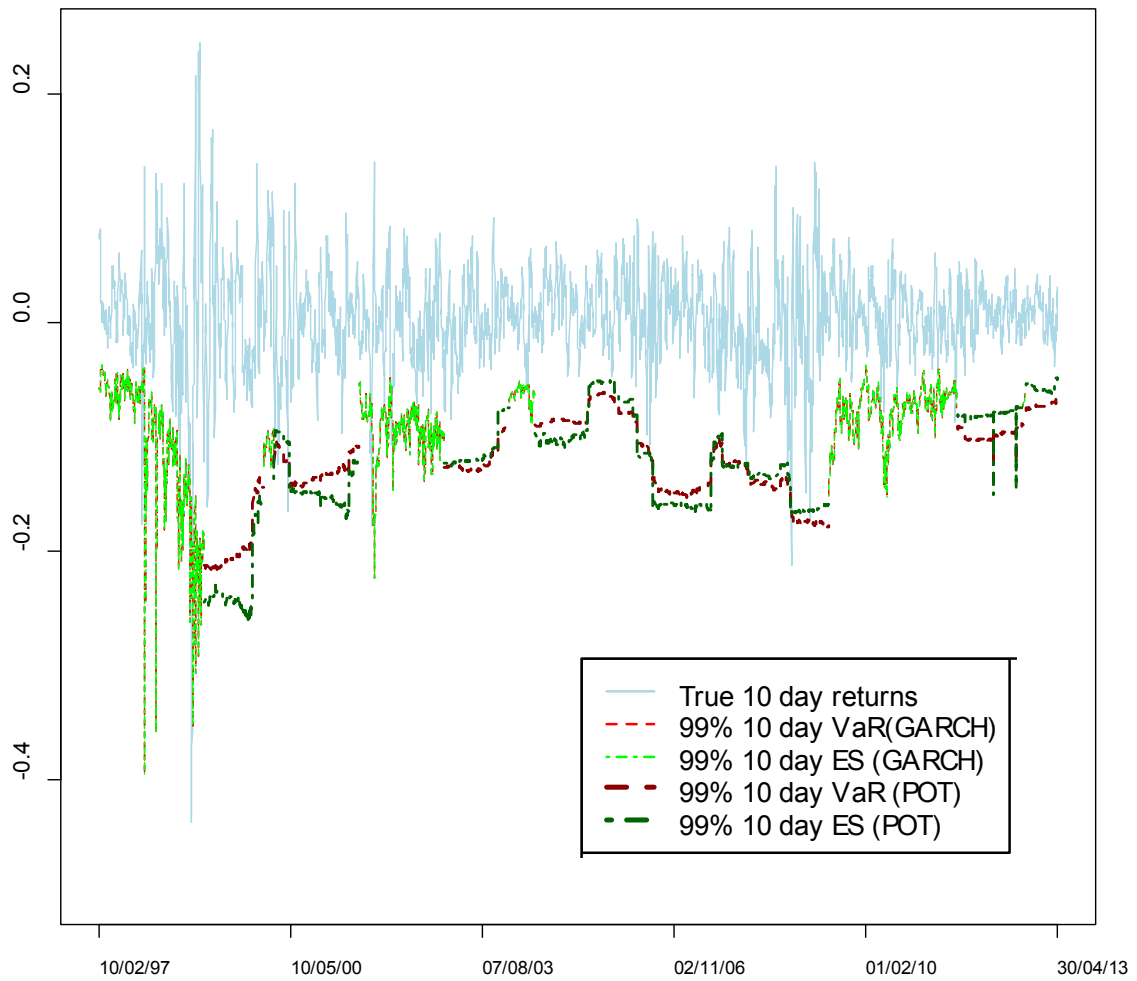


Figure 8.29: Forecasts from the E-GARCH model (Student t innovations) augmented with the POT EVT model (ML estimate of the EVI).

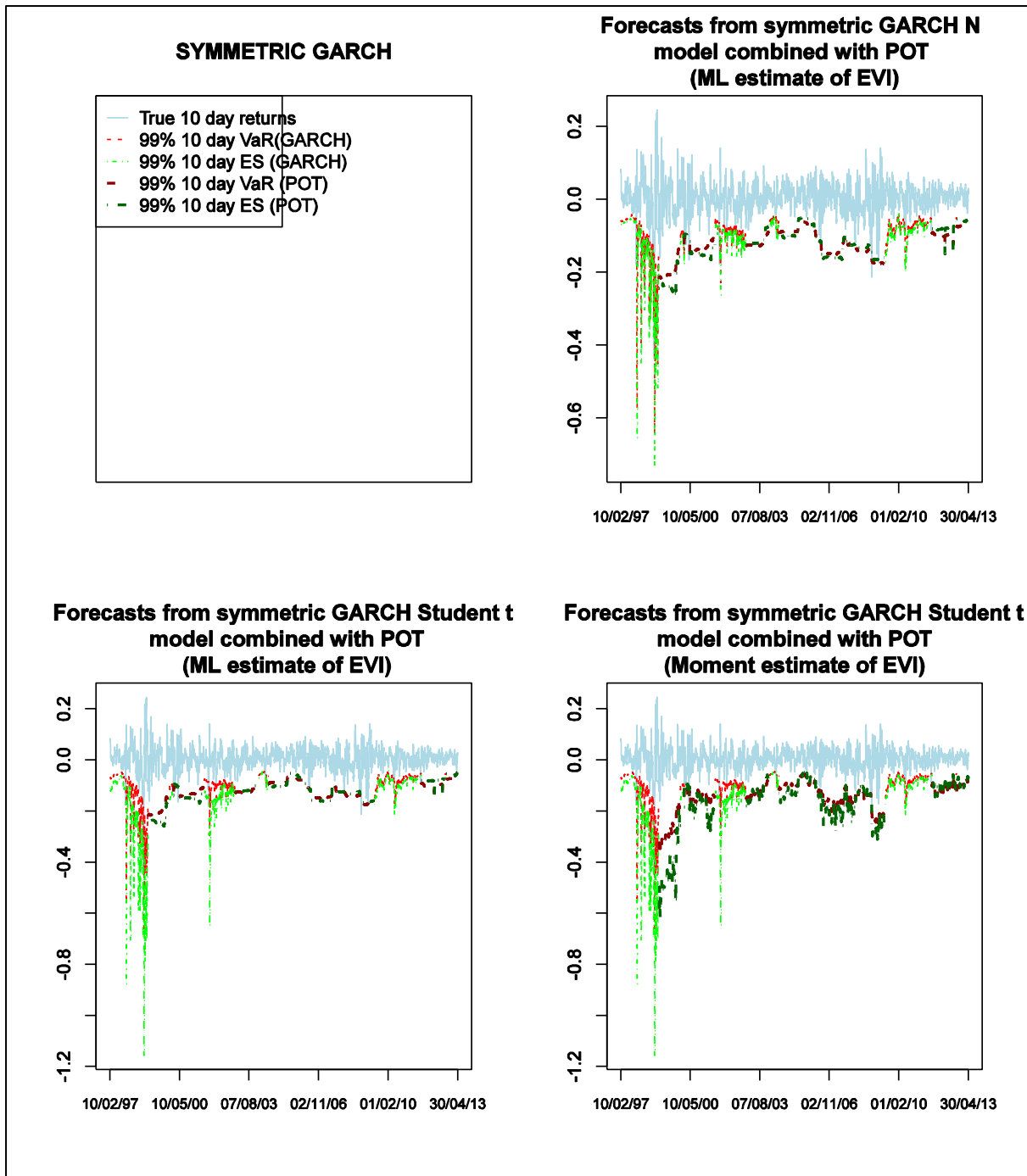


Figure 8.30: Symmetric GARCH models augmented with EVT models.

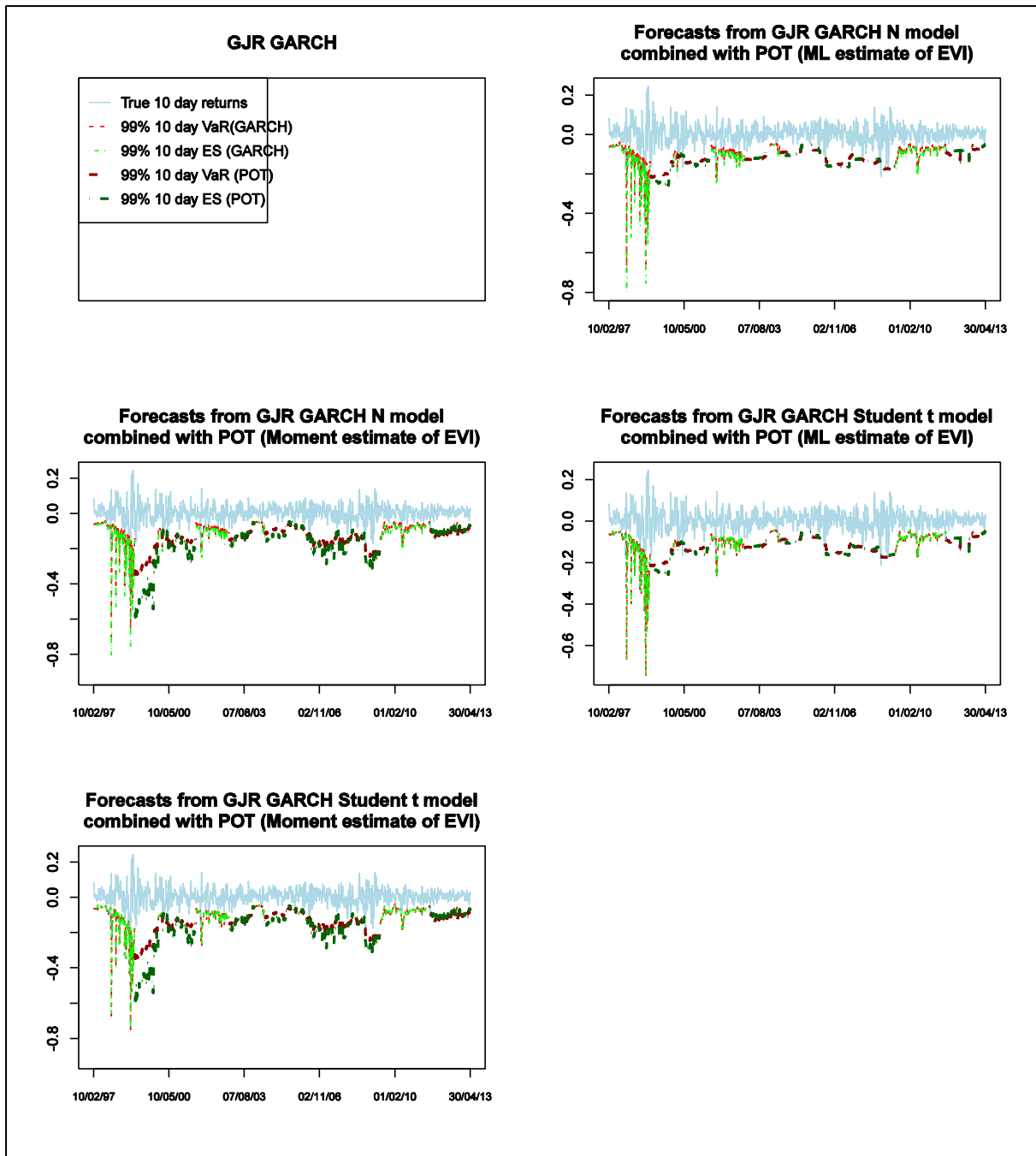


Figure 8.31: Symmetric GJR-GARCH models augmented with EVT models.

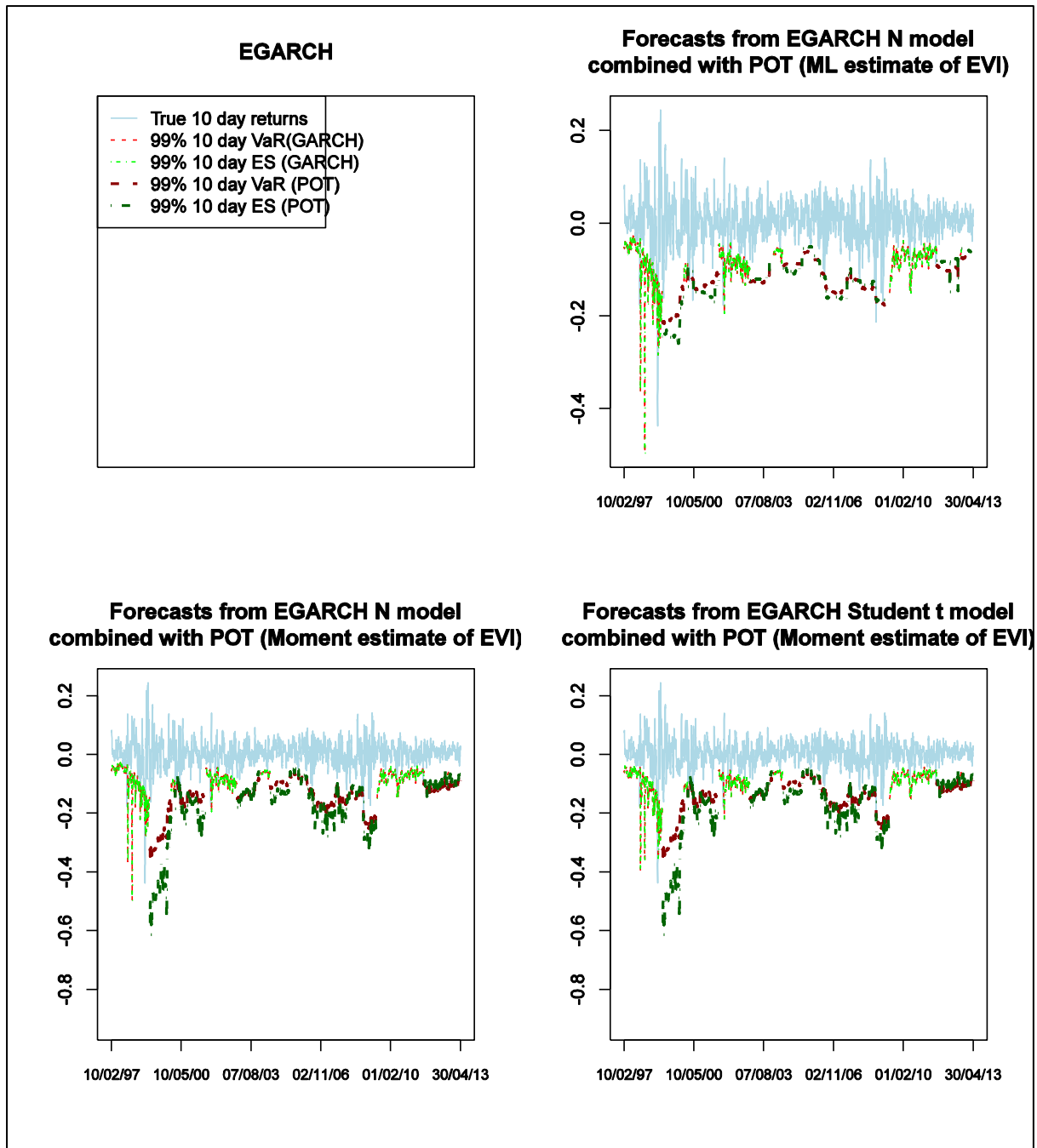


Figure 8.32: E-GARCH models augmented with EVT models.

The violations for the combined models are summarised in Table 8.8 and Figure 8.33 for the VaR forecasts; and in Table 8.9 and Figure 8.34 for the ES forecasts.

Table 8.8: Number of VaR violations for different combinations of GARCH and POT models. (Expected number of violations is 40.54.)

	Number of VaR violations		
	GARCH model	POT model	Total
Symmetric GARCH (N)+POT(ML)	38	4	42
GJR GARCH (N)+POT (ML)	43	4	47
EGARCH (N)+POT(ML)	42	4	46
Symmetric GARCH (N)+POT(Moment)	38	3	41
GJR GARCH (N)+POT (Moment)	43	3	46
EGARCH (N)+POT(Moment)	42	3	45
Symmetric GARCH (t)+POT(ML)	30	4	34
GJR GARCH (t)+POT (ML)	35	4	39
EGARCH (t)+POT(ML)	36	4	40
Symmetric GARCH (t)+POT(Moment)	30	3	33
GJR GARCH (t)+POT (Moment)	35	3	38
EGARCH (t)+POT(Moment)	36	3	39

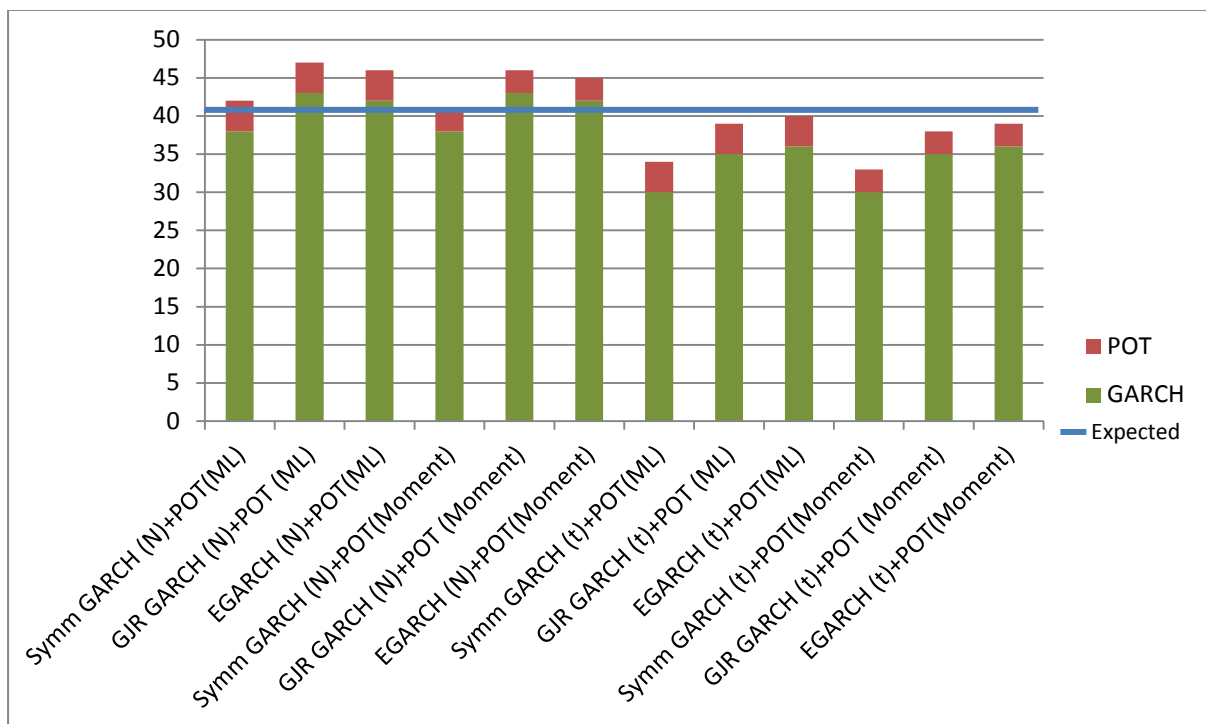


Figure 8.33: Bar chart of number of VaR violations for different combinations of GARCH and POT models (entire data series). (Expected number of violations is 40.45.)

Table 8.9: Number of ES violations for different combinations of GARCH and POT models.

(Expected number of violations is 40.54.)

	Number of ES violations		
	GARCH model	POT model	Total
Symm GARCH (N)+POT(ML)	25	6	31
GJR GARCH (N)+POT (ML)	32	6	38
EGARCH (N)+POT(ML)	42	6	48
Symm GARCH (N)+POT(Moment)	28	5	33
GJR GARCH (N)+POT (Moment)	32	5	37
EGARCH (N)+POT(Moment)	42	5	47
Symm GARCH (t)+POT(ML)	13	6	19
GJR GARCH (t)+POT (ML)	35	6	41
EGARCH (t)+POT(ML)	36	6	42
Symm GARCH (t)+POT(Moment)	13	5	18
GJR GARCH (t)+POT (Moment)	35	5	40
EGARCH (t)+POT(Moment)	36	5	41

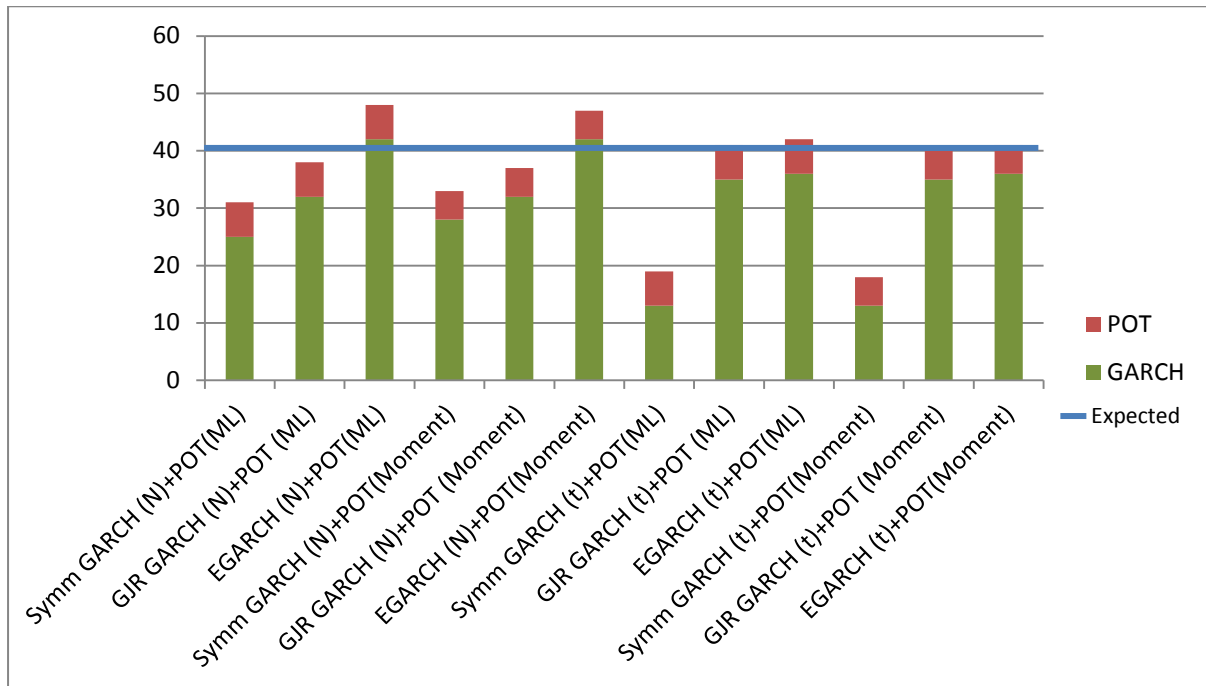


Figure 8.34: Bar chart of number of ES violations for different combinations of GARCH and POT models (entire data series). (Expected number of violations is 40.45.)

Overall, the models perform reasonably well based on the observed number of violations (the expected number is 40.54). Most of the violations stem from the GARCH models. This is

because of the choice of the empirical scaling laws for the POT estimates. This scaling law has been seen to be very conservative.

The two best performing models based on the number of VaR violations (Table 8.8) are:

1. The symmetric GARCH model with the normal innovations combined with the POT model using the ML estimate of the EVI. The number of violations is 42, compared to the expected number of 40.54.
2. The symmetric GARCH model with the Student t innovations combined with the POT model using the Moment estimate of the EVI. The number of violations is 41, compared to the expected number of 40.54.

It is interesting to compare these two best performing models with the i.i.d. models in Section 8.2 and the EVT models in Section 8.4.2.2. The i.i.d. normal model yielded 51 VaR violations and the i.i.d. Student t model yielded 43 (refer Table 8.1). The best performing EVT model in Table 8.7 is the POT model with the ML estimate of the EVI and using the square root of time scaling law. This model gave 42 violations.

The procedure suggested in this section of augmenting the GARCH models with EVT forecasts therefore provides very good results compared to previous procedures. It performs better than the i.i.d. models and about the same as the best performing EVT model. It improves markedly upon the other EVT models in Table 8.7.

8.5. Two step procedure

The two step procedure described in section 5.1 is now applied to obtain ten day 99% VaR estimates. Filtering is performed with the i.i.d. normal and Student t models separately. These models produce residuals that can be used in the EVT procedures to calculate new VaR forecasts. The ML and moment estimators are considered, as well as the three scaling rules discussed in section 4.8.

The pre-filtering with the i.i.d models can be done over the entire data series. However, it is also of interest to pre-filter with a GARCH model. The chosen GARCH model is the symmetric GARCH model with Student t innovations, since this model was one of the best performing models in previous analyses. In this case, filtering cannot be done over the entire data series.

The GARCH models only converge over certain periods. The reasons therefore were outlined in Chapter 6. The main problem here is the presence of extreme observations (outliers) that impede convergence. These periods are summarised in Table 8.10 and will be handled separately. This is mainly for the purpose of the empirical scaling rule, so that the distribution in each segment can be individually used to determine the percentiles. Note also that only periods 1, 3 and 5 are long enough to be used for rolling 250 day estimates.

Table 8.10: Index of values over which all the GARCH models converge:

	Observation Number	
	Start	End
Period 1	1	444
Period 2	695	740
Period 3	1100	1460
Period 4	1736	1842
Period 5	3086	3629
Period 6	3907	3917

8.5.1. Filter with the i.i.d. normal model

The first filtering model considered is the i.i.d. normal model. The ML and Moment estimators of the EVI are used. In each instance, all the three scaling rules are applied. The results for the different models are summarised in Figure 8.35.

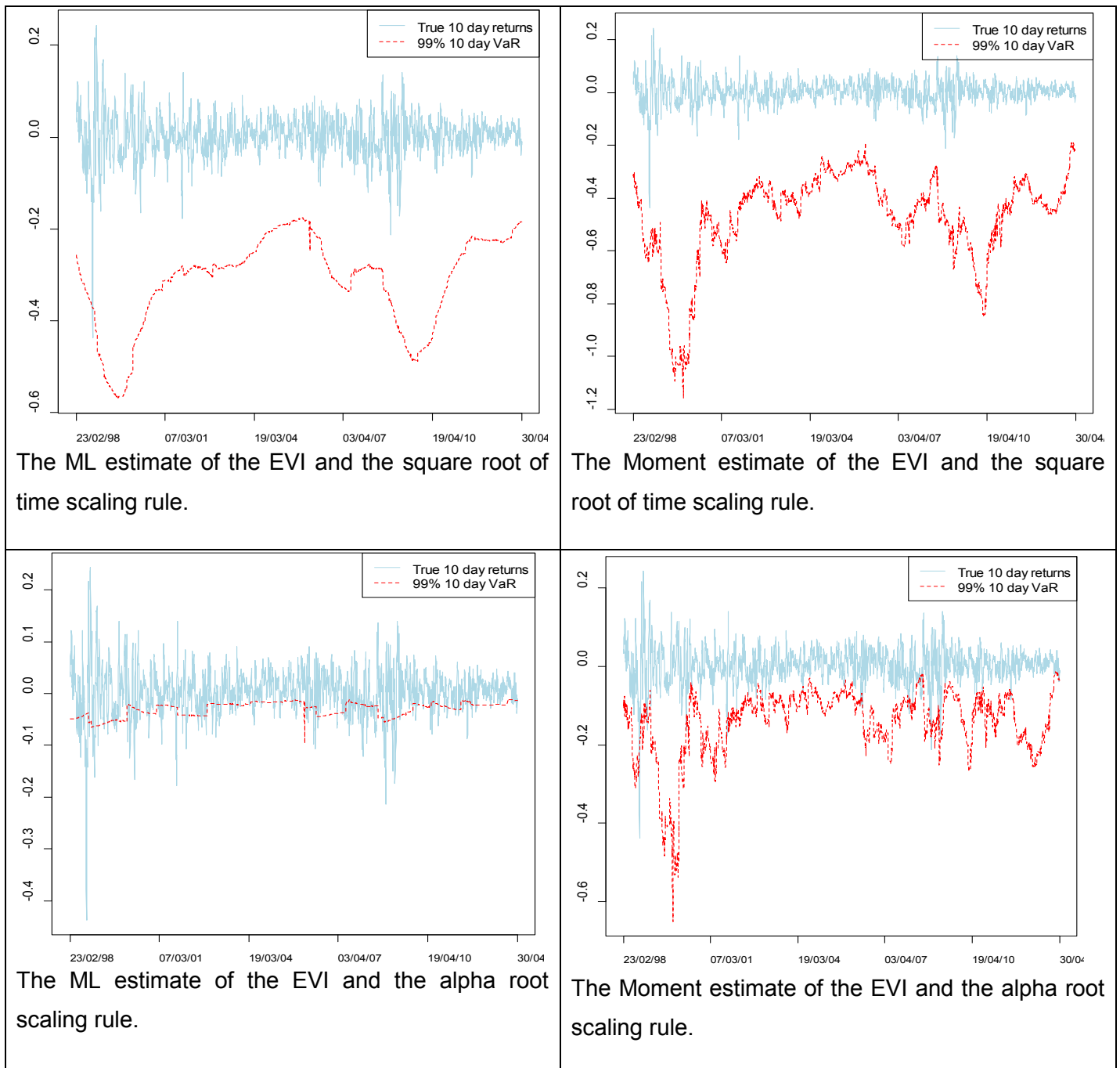


Figure 8.35: Ten day VaR forecasts from the POT method with different estimators of the EVI and different scaling rules, pre-filtered using the i.i.d. normal model.

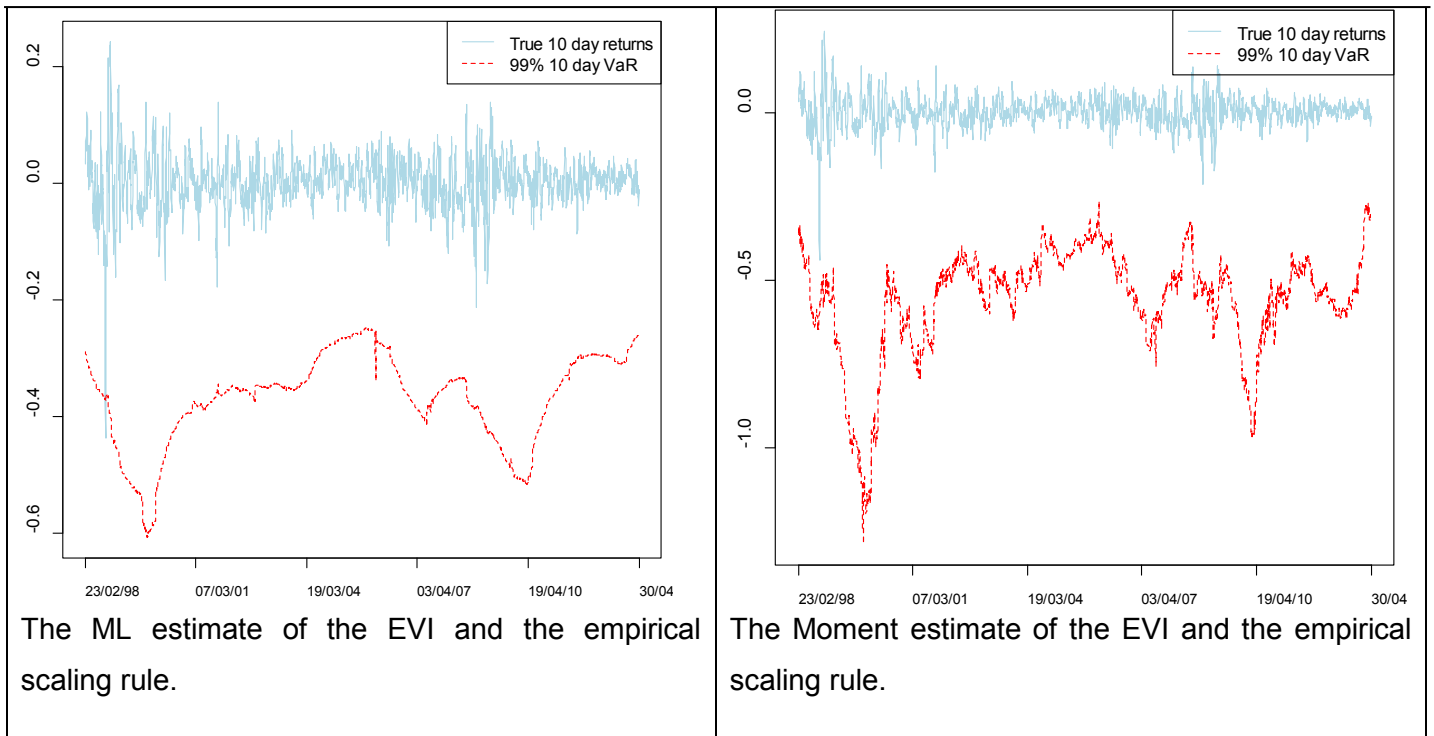


Figure 8.35 (continued): Ten day VaR forecasts from the POT method with different estimators of the EVI and different scaling rules, pre-filtered using the i.i.d. normal model.

8.5.2. Filter with the i.i.d. Student t model

The second filtering model considered is the i.i.d. Student t model. The ML and Moment estimators of the EVI are used, with all the three scaling rules. The results for the different models are summarised in Figure 8.36.

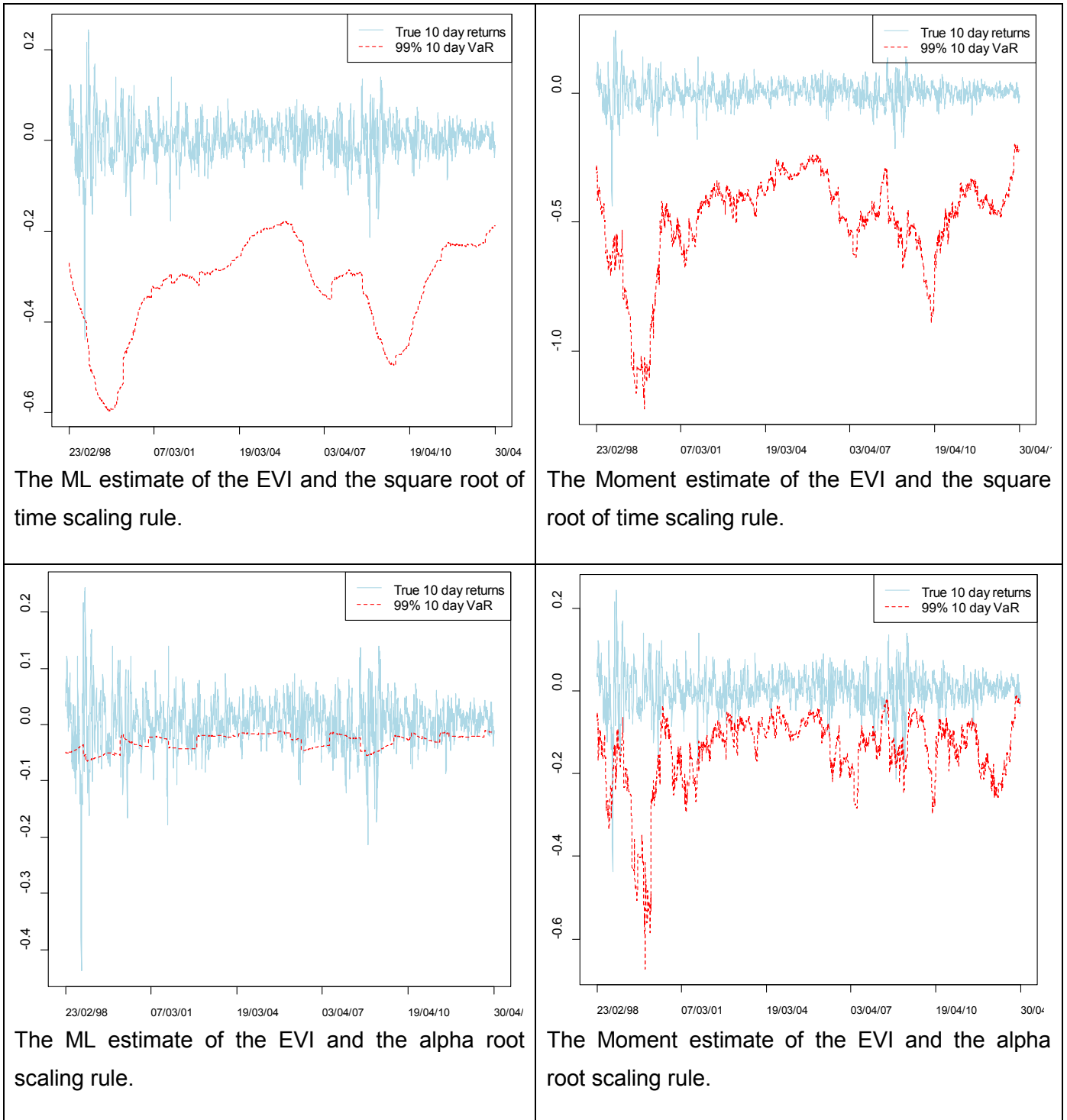


Figure 8.36: Ten day VaR forecasts from the POT method with different estimators of the EVI and different scaling rules, pre-filtered using the i.i.d. Student t model.

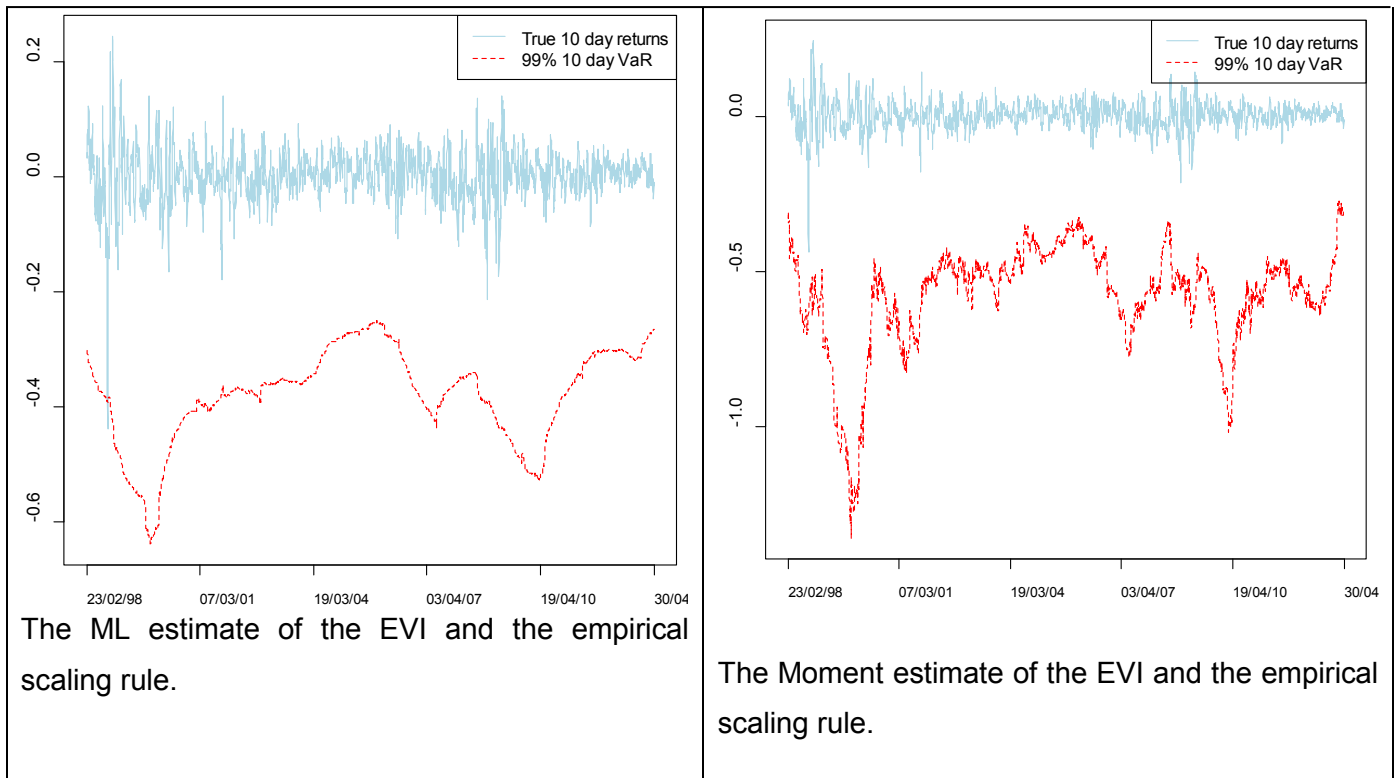


Figure 8.36 (continued): Ten day VaR forecasts from the POT method with different estimators of the EVI and different scaling rules, pre-filtered using the i.i.d. Student t model.

The violations from the two different filtering methods in section 8.5.1 (the i.i.d. normal model) and section 8.5.2 (the i.i.d. Student t model) are summarised in Table 8.11 and Figure 8.37.

Table 8.11: Number of VaR violations from different two step procedures (expected number of violations is 37.95)

		Scaling Rule		
Model for pre-filtering	POT estimate	Square Root	Alpha Root	Empirical
I.i.d. Normal	ML	4	698	5
	Moment	0	58	0
I.i.d. Student t	ML	2	708	2
	Moment	0	45	0

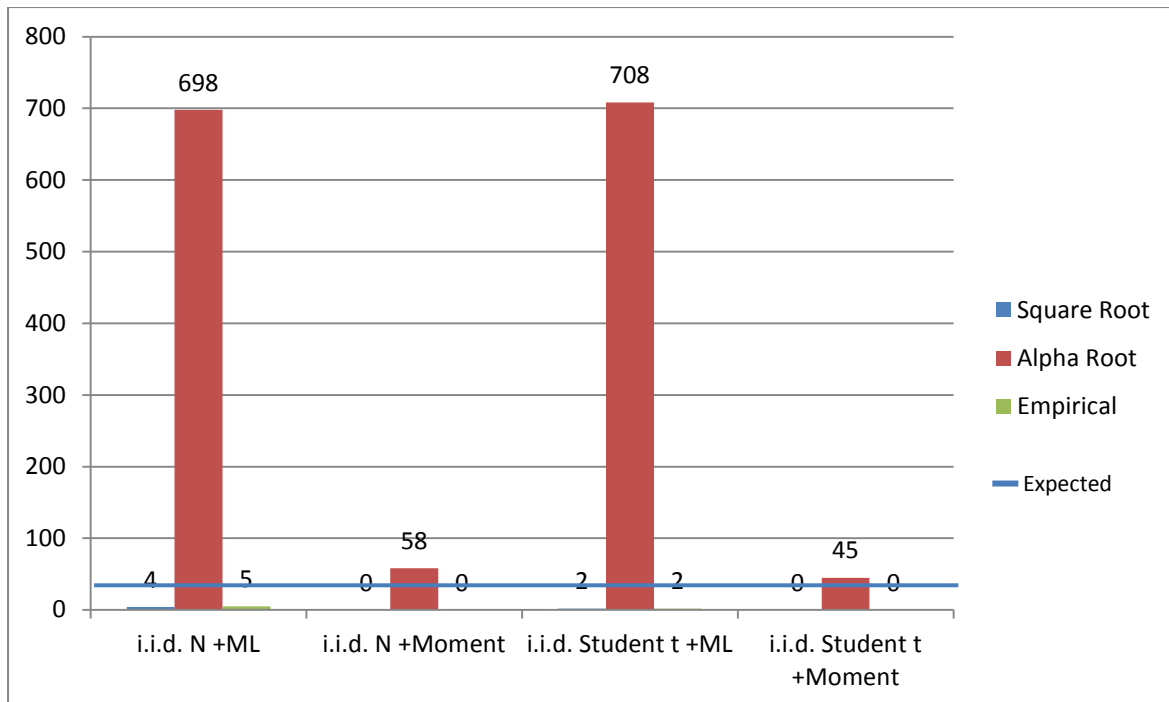


Figure 8.37: Bar chart of number of VaR violations from different two step procedures (expected number of violations is 37.95).

The models do not perform well, since the actual number of violations is in each case very far from the expected number of 37.95. The alpha root scaling rule provides too many violations, especially so for the ML estimate of the EVI. The best forecasts seem to be achieved by using the alpha root scaling rule together with the Moment estimate of the EVI and pre-filtering with the i.i.d. Student t model. For the other models, pre-filtering does not seem useful.

8.5.3. Filter with the symmetric GARCH Student t model

The last filtering method considered, is that using the symmetric GARCH student t model. Figures 8.38, 8.39 and 8.40 give the estimation results for the square root of time scaling rule, alpha root scaling rule, and empirical scaling rules respectively. Both the ML and Moment estimates of the EVI are considered. The plots were only performed for the three feasible periods identified at the start of section 8.5.

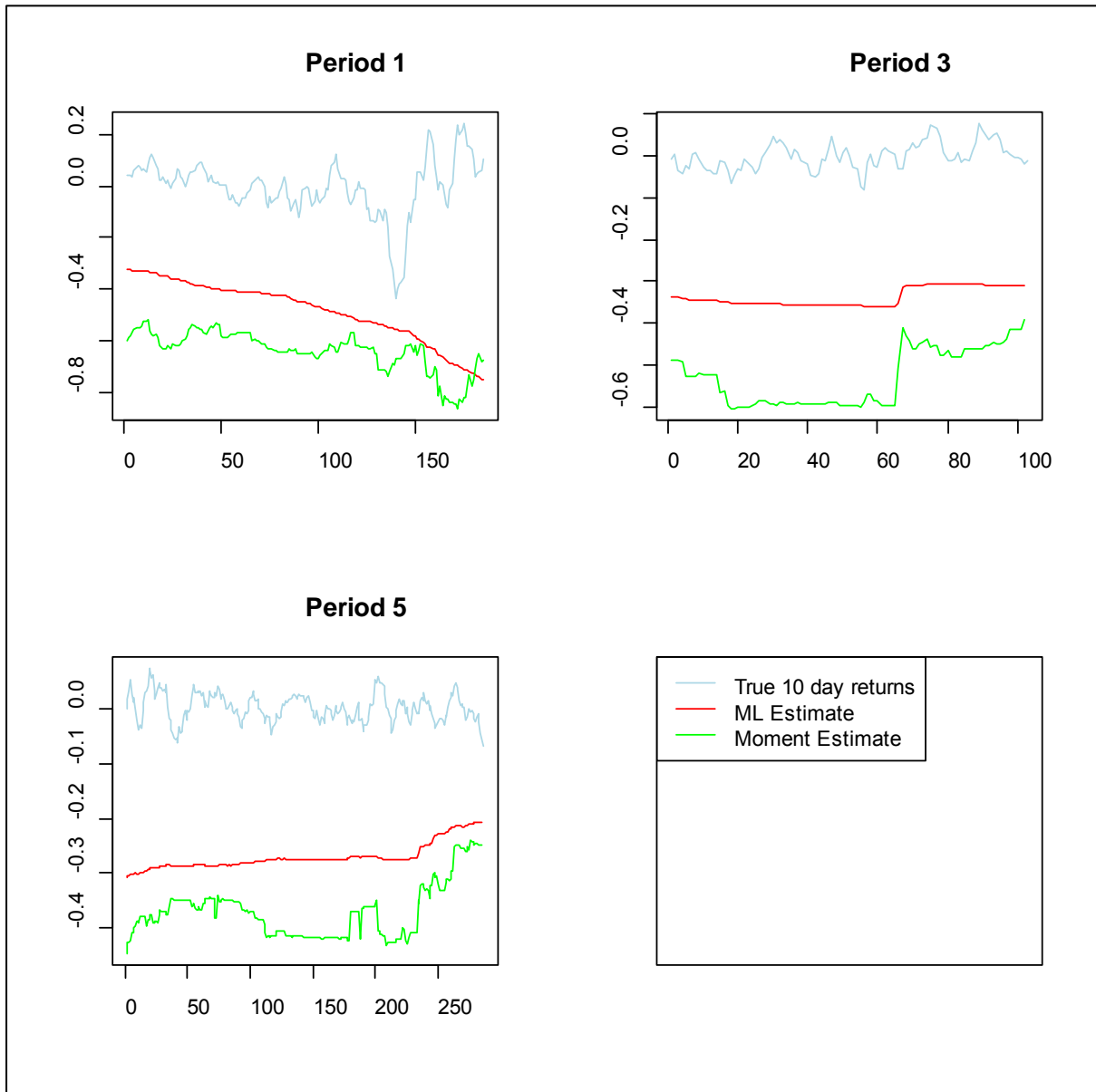


Figure 8.38: Ten day VaR forecasts from the POT method with a Moment and an ML estimate of the EVI, pre-filtered using the symmetric GARCH Student t model and applying the square root of time scaling rule.

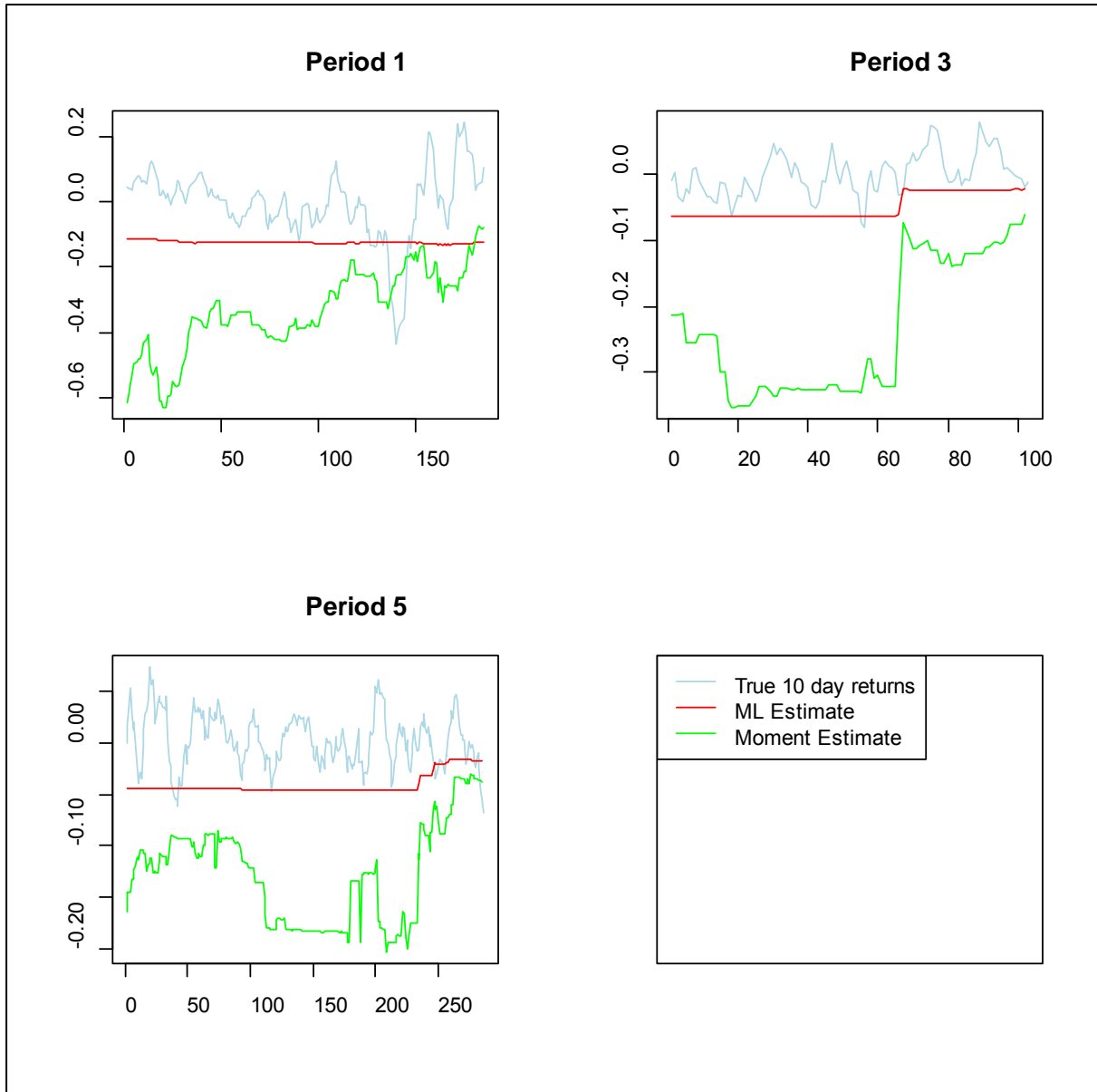


Figure 8.39: Ten day VaR forecasts from the POT method with a Moment and an ML estimate of the EVI, pre-filtered using the symmetric GARCH Student t model and applying the alpha root of time scaling rule.

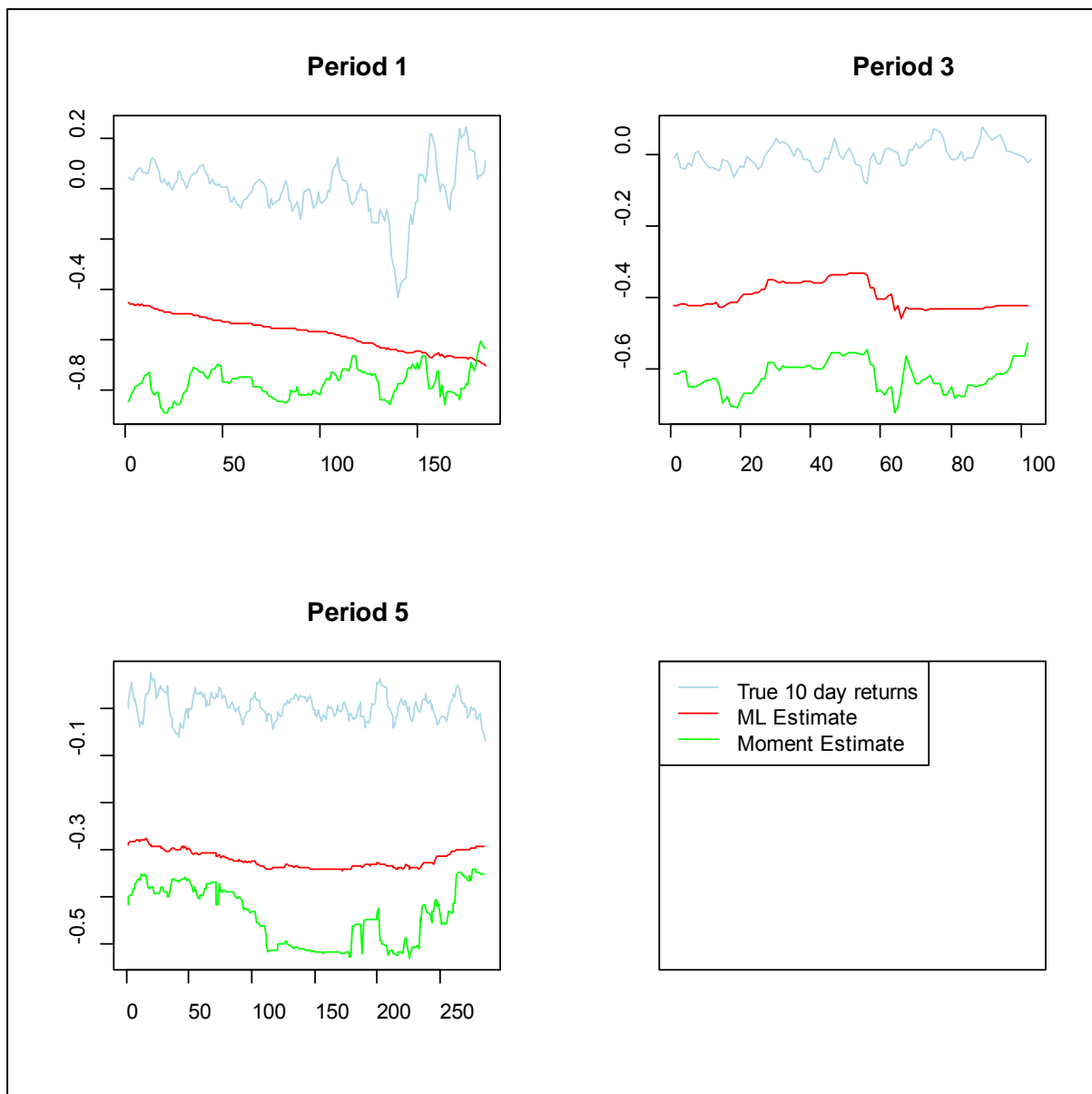


Figure 8.40: Ten day VaR forecasts from the POT method with a Moment and an ML estimate of the EVI, pre-filtered using the symmetric GARCH Student t model and applying the empirical scaling rule.

From Figures 8.38 to 8.40 it can be concluded that pre-filtering with the symmetric GARCH model leads to very conservative VaR forecasts. The number of violations is all zero, except for the ML estimate of the EVI combined with the alpha root scaling rule. This is in agreement with earlier results that both the ML estimate and the alpha root scaling rule seem to provide too many violations. The Moment estimate is almost always below the ML estimate.

The overall conclusion is that the two-step procedure is not useful when pre-filtering with the symmetric GARCH model. However, it must be remembered that pre-filtering could only be done for periods where the GARCH models converged in the first place. These periods are by nature the more tranquil periods of the time series. It therefore makes intuitive sense that the forecasts would be overly conservative, since an already tranquil time series is filtered thus leading to over-fitting.

9. Conclusion and Open Questions

In this thesis various models that can be used to calculate VaR and ES estimates were investigated. The models were evaluated using the Africa Financials Index (J580) data for the period from 29/01/1996 to 30/04/2013. The focus was on estimation of periods containing extreme returns. This is because extreme returns are what constitute market risk and it is important that the forecasting methods incorporate these as accurately as possible.

The aim of the thesis was to explore various forecasting models by a practical application. The different models were compared in terms of the number of VaR and ES violations. Limitations of models were identified and alternative strategies were suggested.

Traditional i.i.d. models and GARCH volatility models were considered. These perform well during relatively tranquil periods containing few extreme return observations. However, during periods of extreme returns, the GARCH models fail to converge. To address this issue of convergence, EVT models were considered that allow especially for the modelling of extreme returns. The EVT models and the POT method were used to obtain forecasts and different scaling methods from one day to ten day estimates. As a novelty, augmentation of the GARCH models with EVT models were investigated and found to perform well compared to previous methods. Application of EVT on returns pre-filtered by selected i.i.d. and GARCH models, as was suggested by McNeil (1999), was also considered.

The five classes of models considered are:

1. Normal and Student t i.i.d. models
2. GARCH stochastic volatility models
3. EVT models
4. GARCH models augmented with EVT models
5. EVT models on pre-filtered GARCH return data.

A selection of results for the VaR forecasts and the ES forecasts for model classes 1 to 4 are presented in Tables 9.1 and 9.2. The two top performing models in each class are given. Performance was evaluated by the closeness of the actual number of violations to the expected number of violations. (The EVT models on the pre-filtered GARCH returns performed so poorly, that the results are not repeated here.)

Table 9.1: Summary of results of VaR forecasts from various models

Model	Number of violations		Percentage above or below expected
	Actual	Expected	
I.i.d. Normal	51	40.45	26.08
I.i.d. Student t	43	40.45	6.30
Symmetric GARCH (t)	30	15.14	98.15
GJR-GARCH (t)	35	15.14	131.18
POT EVT (ML estimate of EVI, Square root scaling)	42	40.45	3.83
POT EVT (Moment estimate of EVI, Square root scaling)	29	40.45	-28.31
Symmetric GARCH (N) + POT (Moment estimate, empirical scaling rule)	41	40.45	1.36
E-GARCH (t) + POT (ML estimate, empirical scaling rule)	40	40.45	-1.11

Table 9.2: Summary of results of ES forecasts from various models

Model	Number of violations		Percentage above or below expected
	Actual	Expected	
I.i.d. Normal	35	40.45	-13.47
I.i.d. Student t	32	40.45	-20.89
Symmetric GARCH (N)	28	15.14	84.94
Symmetric GARCH (t)	13	15.14	-14.13
POT EVT (ML estimate of EVI, Square root scaling)	49	40.45	21.14
POT EVT (Moment estimate of EVI, Square root scaling)	28	40.45	-30.78
GJR-GARCH (t) + POT (ML estimate, empirical scaling rule)	41	40.45	1.36
GJR-GARCH (t) + POT (Moment estimate, empirical scaling rule)	41	40.45	1.36
GJR-GARCH (t) + POT (Moment estimate, empirical scaling rule)	40	40.45	-1.11

The overall conclusion is that none of the models explored is universally optimal. Some of the models perform well, but other data sets should also be investigated to generalise the conclusions. The key findings in this study are summarised in the following points:

- The best performing models are those that augment the GARCH models with the EVT forecasts. The E-GARCH with student t innovations augmented with the POT ML estimate of the EVI performs the best among all the models for VaR (Table 9.1). The GJR-GARCH with student t innovations augmented with the POT Moment estimate is optimal for the ES models. Both yield 40 violations, which is close to the

expected number of 40.45. The other top-performing models in this class yield 41 violations. Allowing for random variation, their performance is equally good.

- This highlights the importance of combining different models, especially in areas where convergence is a problem due to extreme returns. The models with the Student t innovations perform better than those with the normal innovations. It is therefore helpful to allow for the fat-tailed nature of the return distribution.
- The i.i.d. models perform reasonably well, considering they are the simplest of the models. The i.i.d. models with Student t innovations are good (and simple) alternatives to more complex models.
- The GARCH models on their own perform very badly. The number of violations is in general far above the expected number of 15.14, even for the best performing models. Again, the Student t models slightly outperform the normal models. The symmetric GARCH models tend to yield better results than the E-GARCH and GJR-GARCH models. However, it should be remembered that the GARCH models are only fitted over periods where convergence is possible. These periods excluded extreme returns. This may be the reason why the symmetric models outperform the other, since the underlying distribution is not overly skewed.
- The EVT POT models differ in performance according to what scaling rule and estimate for the EVI is used. In general, the moment estimates yield fewer violations than the ML estimates. The alpha root scaling rule yields by far the most violations and performs the poorest, whereas the empirical scaling rule yields the least number of violations. The square root of time scaling rule comes closest to the expected number of violations for both the VaR and ES estimates.
- The two-step procedure suggested by McNeil (1999) of first pre-filtering the return data with a volatility model and then applying EVT to the residuals, is not successful for this data set. The number of violations is mostly zero, or close to zero, except for

the alpha root scaling rule, which has been seen to be overly conservative. This type of VaR forecast is not insightful, since it is no better than assuming an arbitrarily large negative VaR value.

- The ES estimates yield consistently lower numbers of violations than the VaR counterparts, except for the EVT models where the order is sometimes reversed. The ES forecast tend to fall below the VaR forecasts in the plots of the estimates. This confirms that ES estimates are more conservative than VaR estimates, a reason why they are sometimes preferred in practice.

It is important to understand the limitations of each model. Especially those of the GARCH models that are unable to accommodate extreme returns. The extreme returns result in failure of convergence of the GARCH models over parts of the sample period. This fact is not properly highlighted in the current literature. The main limitation of the EVT models is that the results are only asymptotically correct. It has been applied here to small samples and therefore the results should be interpreted with care. It is equally important to realise that alternatives are available and that combinations of models should be considered for optimal forecasting. In this study, the best alternative is to fill the gaps from the GARCH forecasts with EVT forecasts. In practice, the chosen modelling method always need to entail critical appraisal of the data at hand, allowable assumptions and objectives that are set.

Although the aims of this study were achieved, there are still a number of open questions that provide scope for further research. This study only considered a single data set. More extensive simulation studies need to be performed in order to generalise the findings of this study. The question of how to scale EVT estimates from one day to ten day estimates is not well researched. Better scaling methods may well be found. Lastly, different i.i.d. models can be considered. The Student t model performs reasonably well, suggesting that other fat-tailed distributions may be sensible to consider and may even provide better results. VaR and ES estimation as well as the modelling of volatility remain active and relevant research fields.

Appendix:

R Code for Chapters 7 and 8

```

=====
#CHAPTER 7: THE DATA
=====

#Figure 7.1:
J580.plot<-read.table(file="clipboard") #the index values
plot(J580.plot[,2],ty="l",xaxt="n",ylab="",xlab="")
n.dates.1<-nrow(J580.plot)
spacings<-c(0.2,0.4,0.6,0.8,1)
ax.vals.1<-c(1,floor(n.dates.1*spacings))
par(las=2)
axis(side=1,at=ax.vals.1,labels=J580.plot[,1][ax.vals.1],cex.axis=0.7)

#Figure 7.2:
#data.all: All the data points (all the log returns)
dates.logs<-read.table(file="clipboard") #dates corresponding to log returns
plot(x=1:length(data.all),y=data.all,ty="l",xlab="",ylab="",main="Log Returns of Africa
Financials Index",xaxt="n")
n.dates.2<-length(data.all)
spacings<-c(0.2,0.4,0.6,0.8,1)
ax.vals.2<-c(1,floor(n.dates.2*spacings))
par(las=2)
axis(side=1,at=ax.vals.2,labels=dates.logs[ax.vals.2,],cex.axis=0.7)

#Figure 7.3:
hist(data.all,xlab="Log return",ylab="Frequency",main="Histogram of Log Returns",breaks=16)

#Figure 7.4:
#Plot of squared log returns:
plot(x=1:length(data.all),y=(data.all)^2,ty="l",xlab="",ylab="",xaxt="n")
axis(side=1,at=ax.vals.2,labels=dates.logs[ax.vals.2,],cex.axis=0.7)

```

```

#Table 7.1 (descriptive statistics):

library(tseries)

library(TSA)

mean<-mean(data.all)

median<-median(data.all)

std.dev<-sd(data.all)

max<-max(data.all)

min<-min(data.all)

inter.quartile.range<-quantile(data.all,0.75)-quantile(data.all,0.25)

excess.kurt<-kurtosis(data.all)

skew<-skewness(data.all)

adf.test(data.all)

jarque.bera.test(data.all)

plot(x=1:length(data.all),y=(data.all)^2,ty="l",xlab="",ylab="",main="Squared Log Returns of
Africa Financials Index")

acf(data.all, main="ACF for log returns")

acf((data.all)^2,main="ACF for squared log returns")

acf(abs(data.all),main="ACF for absolute log returns")

#To test for asymmetry effects:

cor((data.all^2)[-1],data.all[-(length(data.all))])

#=====

#CHAPTER 8: ANALYSIS OF THE AFRICA FINANCIALS INDEX DATA

#=====

#=====

#SECTION 8.2.: INDEPENDENT IDENTICALLY DISTRIBUTED MODELS

#=====

J580<-read.table(file="clipboard")

data<-J580

#This is the log returns of the Africa Financial index from 29/01/96 to 30/04/13

#Number of observations=4313

m<-250 #The length of the moving data window

#This window can start at observation=1, but must end at observation=4054,

#since this is the last 250 day period that will still enable a comparison

```

```

#with 10 days ahead.
n<-4054 #The number of 10 day ahead forecasts that will be made

#=====
#N IID MODEL
#=====

iid.n.vol<-rep(0,n) #the forecasted yearly volatility for 10 days ahead
iid.n.VaR<-rep(0,n) #the iid VaR forecasts
iid.n.ES<-rep(0,n) #the iid ES forecasts

for (i in 1:n){
  sdev<-sd(data[i:(i+250-1),])
  iid.n.vol[i]<-sdev*sqrt(10)
  iid.n.VaR[i]<--sdev*qnorm(0.99)*sqrt(10)
  iid.n.ES[i]<--(1/0.01)*dnorm(qnorm(0.01))*sdev*sqrt(10)
}

iid.n.vol.save<-iid.n.vol
iid.n.VaR.save<-iid.n.VaR
iid.n.ES.save<-iid.n.ES

#Figure 8.1:
true<-data[260:4313,] #daily returns: for the plot we need 10 day returns
#true.10<-read.table(file="clipboard") #10 day log returns

ylim<-
c(min(iid.n.vol.save,iid.n.VaR.save,iid.n.ES.save,true.10),max(iid.n.vol.save,iid.n.VaR.save,iid.n.ES.save,true.10))

plot(true.10,lty=1, main="IID N model",ty="l",col="light
blue",ylim=ylim,xlab="",ylab="",xaxt="n")

lines(iid.n.vol.save,lty=3)
lines(iid.n.VaR.save,lty=2,col="red")
lines(iid.n.ES.save,lty=4,col="green")

legend(x="bottomright",legend=c("True 10 day log returns","99% 10 day VaR
forecast","Volatility forecast","99% 10 day ES"),
lty=1:4,col=c("light blue","red","black","green"))

#dates.log.10<-read.table(file="clipboard")

ax.vals.3<-c(1,floor(length(true.10)*spacings))
axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.7)

```

```

iid.n.VaR.x<-sum(true.10<iid.n.VaR)
iid.n.ES.x<-sum(true.10<iid.n.ES)

#Figure 8.5:
#plot only overlapping areas:
plot(true.10,lty=1, main="IID N model",ty="l",col="light
blue",ylim=ylim,xlab="",ylab="",xaxt="n")

iid.n.vol[indices.empty]<-" "
lines(iid.n.vol,lty=3)

iid.n.VaR[indices.empty]<-" "
lines(iid.n.VaR,lty=2,col="red")

iid.n.ES[indices.empty]<-" "
lines(iid.n.ES,lty=4,col="green")

legend(x="bottomright",legend=c("True 10 day log returns","99% 10 day VaR
forecast","Volatility forecast","99% 10 day ES"),
lty=1:4,col=c("light blue","red","black","green"))

axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.8)

#####
#STUDENT t IID MODEL
#####

#To get df using ML methods:
df.ML<-rep(0,n)
for (i in 1:n){
  dats<-data[i:(i+250-1),]
  sdev<-sd(dats)
  mean<-mean(dats)
  dat.t<-(dats-mean)/sdev #standardise data for standardised t

  t.mle<-function(dat=dat.t){
    t.lik<-function(nu,y){
      #y will be the data
      logl<-sum(log(dt(x=y,df=nu)))
      return(-logl) #return negative because of optim()
    }

    results<-optim(par=4,fn=t.lik,y=dat,method="BFGS")
  }
}

```

```

    print(results)
  }
  df.ML[i]<-t.mle()[[1]]
}

iid.t.vol<-rep(0,n) #same as for N?
iid.t.VaR<-rep(0,n)
iid.t.ES<-rep(0,n)

for (i in 1:n){
  sdev<-sd(data[i:(i+250-1),])
  iid.t.vol[i]<-sdev*sqrt(10)
  df<-df.ML[i]
  iid.t.VaR[i]<--sqrt((1/df)*(df-2))*qt(p=0.99,df=df)*sdev*sqrt(10)
  iid.t.ES[i]<--(1/0.01)*(1/(df-1))*(df-
2+(qt(p=0.01,df=df))^2)*dt(qt(p=0.01,df=df),df=df)*sdev*sqrt(10)
}

iid.t.vol.save<-iid.t.vol
iid.t.VaR.save<-iid.t.VaR
iid.t.ES.save<-iid.t.ES

```

Figure 8.2:

```

true<-data[260:4313,]

ylim<-
c(min(iid.t.vol.save,iid.t.VaR.save,iid.t.ES.save,true.10),max(iid.t.vol.save,iid.t.VaR.save,iid.t.ES.save,true.10))

plot(true.10,lty=1, main="IID t model",ty="l",col="light
blue",ylim=ylim,xlab="",ylab="",xaxt="n")

lines(iid.t.vol.save,lty=3)

lines(iid.t.VaR.save,lty=2,col="red")

lines(iid.t.ES.save,lty=4,col="green")

legend(x="bottomright",legend=c("True 10 day log returns","99% 10 day VaR
forecast","Volatility forecast","99% 10 day ES"),

lty=1:4,col=c("light blue","red","black","green"))

axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.7)

iid.t.VaR.x<-sum(true.10<iid.t.VaR)

iid.t.ES.x<-sum(true.10<iid.t.ES)

```


Figure 8.5:

```

true<-data[260:4313,]

ylim<-
c(min(iid.t.vol.save,iid.t.VaR.save,iid.t.ES.save,true.10),max(iid.t.vol.save,iid.t.VaR.save,i
iid.t.ES.save,true.10))

plot(true.10,lty=1, main="IID t model",ty="l",col="light
blue",ylim=ylim,xlab="",ylab="",xaxt="n")

iid.t.vol[indices.empty]<-" "
lines(iid.t.vol,lty=3)

iid.t.VaR[indices.empty]<-" "
lines(iid.t.VaR,lty=2,col="red")

iid.t.ES[indices.empty]<-" "
lines(iid.t.ES,lty=4,col="green")

legend(x="bottomright",legend=c("True 10 day log returns","99% 10 day VaR
forecast","Volatility forecast","99% 10 day ES"),
lty=1:4,col=c("light blue","red","black","green"))

axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.8)

plot(df.ML,ty="l",main="Evolution of df in i.i.d. student t model")

#####
#SECTION 8.3.: GARCH MODELS
#####

#####
#SECTION 8.3.1.: ESTIMATION RESULTS
#####

#####
#SYMM GARCH N MODEL
#Use the daily log returns to fit the models.
#The 1 day ahead forecasts are then calculated. These are the GARCH term structure.
#####

library(rugarch)

spec.garch.n<-
ugarchspec(variance.model=list(model="sGARCH",garchOrder=c(1,1)),mean.model=list(armaOrder=c(0
,0)),distribution.model="norm")

#remember to change the model and distribution if needed

garch.n.vol<-rep(0,n)

```

```

garch.n.VaR<-rep(0,n)
garch.n.ES<-rep(0,n)
w.n<-rep(0,n) #To extract the fitted parameter values
a.n<-rep(0,n)
b.n<-rep(0,n)

for (i in 1:444){
  fit.garch.n.i<-ugarchfit(spec=spec.garch.n,data=data[i:(i+250-1)],)
  w.n[i]<-coef(fit.garch.n.i)[["omega"]]
  a.n[i]<-coef(fit.garch.n.i)[["alpha"]]
  b.n[i]<-coef(fit.garch.n.i)[["beta1"]]
  forc.garch.n.i<-ugarchforecast(fit.garch.n.i,n.ahead=10)
  sigma.squareds.i<-sigma(forc.garch.n.i)^2
  garch.n.vol[i]<-sqrt(sum(sigma.squareds.i)/10)
  garch.n.VaR[i]<--sqrt(sum(sigma.squareds.i)*(1/10))*qnorm(0.99)
  garch.n.ES[i]<--(1/0.01)*dnorm(qnorm(0.01))*sqrt(sum(sigma.squareds.i)*(1/10))
}

for (i in 695:849){
  fit.garch.n.i<-ugarchfit(spec=spec.garch.n,data=data[i:(i+250-1)],)
  w.n[i]<-coef(fit.garch.n.i)[["omega"]]
  a.n[i]<-coef(fit.garch.n.i)[["alpha"]]
  b.n[i]<-coef(fit.garch.n.i)[["beta1"]]
  forc.garch.n.i<-ugarchforecast(fit.garch.n.i,n.ahead=10)
  sigma.squareds.i<-sigma(forc.garch.n.i)^2
  garch.n.vol[i]<-sqrt(sum(sigma.squareds.i)/10)
  garch.n.VaR[i]<--sqrt(sum(sigma.squareds.i)*(1/10))*qnorm(0.99)
  garch.n.ES[i]<--(1/0.01)*dnorm(qnorm(0.01))*sqrt(sum(sigma.squareds.i)*(1/10))
}

for (i in 1100:1460){
  fit.garch.n.i<-ugarchfit(spec=spec.garch.n,data=data[i:(i+250-1)],)
  w.n[i]<-coef(fit.garch.n.i)[["omega"]]
  a.n[i]<-coef(fit.garch.n.i)[["alpha"]]
  b.n[i]<-coef(fit.garch.n.i)[["beta1"]]
  forc.garch.n.i<-ugarchforecast(fit.garch.n.i,n.ahead=10)
  sigma.squareds.i<-sigma(forc.garch.n.i)^2

```

```

garch.n.vol[i]<-sqrt(sum(sigma.squareds.i)/10)
garch.n.VaR[i]<--sqrt(sum(sigma.squareds.i)*(1/10))*qnorm(0.99)
garch.n.ES[i]<--(1/0.01)*dnorm(qnorm(0.01))*sqrt(sum(sigma.squareds.i)*(1/10))
}

for (i in 1711:2039){
  fit.garch.n.i<-ugarchfit(spec=spec.garch.n,data=data[i:(i+250-1)],)
  w.n[i]<-coef(fit.garch.n.i)[["omega"]]
  a.n[i]<-coef(fit.garch.n.i)[["alpha1"]]
  b.n[i]<-coef(fit.garch.n.i)[["beta1"]]
  forc.garch.n.i<-ugarchforecast(fit.garch.n.i,n.ahead=10)
  sigma.squareds.i<-sigma(forc.garch.n.i)^2
  garch.n.vol[i]<-sqrt(sum(sigma.squareds.i)/10)
  garch.n.VaR[i]<--sqrt(sum(sigma.squareds.i)*(1/10))*qnorm(0.99)
  garch.n.ES[i]<--(1/0.01)*dnorm(qnorm(0.01))*sqrt(sum(sigma.squareds.i)*(1/10))
}

for (i in 2990:n){
  fit.garch.n.i<-ugarchfit(spec=spec.garch.n,data=data[i:(i+250-1)],)
  w.n[i]<-coef(fit.garch.n.i)[["omega"]]
  a.n[i]<-coef(fit.garch.n.i)[["alpha1"]]
  b.n[i]<-coef(fit.garch.n.i)[["beta1"]]
  forc.garch.n.i<-ugarchforecast(fit.garch.n.i,n.ahead=10)
  sigma.squareds.i<-sigma(forc.garch.n.i)^2
  garch.n.vol[i]<-sqrt(sum(sigma.squareds.i)/10)
  garch.n.VaR[i]<--sqrt(sum(sigma.squareds.i)*(1/10))*qnorm(0.99)
  garch.n.ES[i]<--(1/0.01)*dnorm(qnorm(0.01))*sqrt(sum(sigma.squareds.i)*(1/10))
}

#The following are 1 day estimates:
garch.n.vol.save<-garch.n.vol
garch.n.VaR.save<-garch.n.VaR
garch.n.ES.save<-garch.n.ES

#Multiply by square root 10 to obtain 10 day estimates:
garch.n.vol.10<-garch.n.vol.save*sqrt(10)
garch.n.VaR.10<-garch.n.VaR.save*sqrt(10)

```

```

garch.n.ES.10<-garch.n.ES.save*sqrt(10)

garch.n.vol.10.save<-garch.n.vol.10
garch.n.VaR.10.save<-garch.n.VaR.10
garch.n.ES.10.save<-garch.n.ES.10

w.n.save<-w.n
a.n.save<-a.n
b.n.save<-b.n

#Figure 8.4:
true<-data[260:4313,] #Rather plot 10 day returns

ylim<-
c(min(garch.n.vol.10.save,garch.n.VaR.10.save,garch.n.ES.10.save,true.10),max(garch.n.vol.10.s
ave,garch.n.VaR.10.save,garch.n.ES.10.save,true.10))

plot(true.10,lty=1, main="Symmetric GARCH N model",ty="l",col="light
blue",ylim=ylim,xlab="",ylab="",xaxt="n")

garch.n.vol.10[indices.empty]<-" "
lines(garch.n.vol.10,lty=3)

garch.n.VaR.10[indices.empty]<-" "
lines(garch.n.VaR.10,lty=2,col="red")

garch.n.ES.10[indices.empty]<-" "
lines(garch.n.ES.10,lty=4,col="green")

legend(x="bottomright",legend=c("True 10 day log returns","99% 10 day VaR
forecast","Volatility forecast","99% 10 day ES"),
lty=1:4,col=c("light blue","red","black","green"))

axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.7)

par(mfrow=c(3,1))

w.n[indices.empty]<-" "
plot(w.n.save,ty="n",main="Evolution of
omega",xlab="",ylab="",ylim=c(min(w.n.save),max(w.n.save)))

lines(w.n)

a.n[indices.empty]<-" "
plot(a.n.save,ty="n",main="Evolution of alpha",xlab="",ylab="")

lines(a.n)

b.n[indices.empty]<-" "

```

```

plot(b.n.save,ty="n",main="Evolution of beta",xlab="",ylab="")

lines(b.n)

par(mfrow=c(1,1))

#####

#SYMM GARCH t MODEL

#####

spec.garch.t<-
ugarchspec(variance.model=list(model="sGARCH",garchOrder=c(1,1)),mean.model=list(armaOrder=c(0
,0)),distribution.model="std")

garch.t.vol<-rep(0,n)

garch.t.VaR<-rep(0,n)

garch.t.ES<-rep(0,n)

w.t<-rep(0,n) #To extract the fitted parameter values

a.t<-rep(0,n)

b.t<-rep(0,n)

for (i in 1:444){

  fit.garch.t.i<-ugarchfit(spec=spec.garch.t,data=data[i:(i+250-1)],)

  w.t[i]<-coef(fit.garch.t.i)[["omega"]]

  a.t[i]<-coef(fit.garch.t.i)[["alpha1"]]

  b.t[i]<-coef(fit.garch.t.i)[["beta1"]]

  df<-coef(fit.garch.t.i)[["shape"]] #Kan ek hierdie df gebruik en nie weer optimering
loop nie?

  forc.garch.t.i<-ugarchforecast(fit.garch.t.i,n.ahead=10)

  sigma.squareds.i<-sigma(forc.garch.t.i)^2

  garch.t.vol[i]<-sqrt(sum(sigma.squareds.i)/10)

  garch.t.VaR[i]<-sqrt((1/df)*(df-2))*qt(p=0.99,df=df)*sqrt(sum(sigma.squareds.i)/10)

  garch.t.ES[i]<--(1/0.01)*(1/(df-1))*(df-
2+(qt(p=0.01,df=df))^2)*dt(qt(p=0.01,df=df),df=df)*sqrt(sum(sigma.squareds.i)/10)

}

for (i in 695:739){

  fit.garch.t.i<-ugarchfit(spec=spec.garch.t,data=data[i:(i+250-1)],)

  w.t[i]<-coef(fit.garch.t.i)[["omega"]]

  a.t[i]<-coef(fit.garch.t.i)[["alpha1"]]

  b.t[i]<-coef(fit.garch.t.i)[["beta1"]]

```

```

    df<-coef(fit.garch.t.i)[["shape"]] #Kan ek hierdie df gebruik en nie weer optimering
loop nie?

    forc.garch.t.i<-ugarchforecast(fit.garch.t.i,n.ahead=10)

    sigma.squareds.i<-sigma(forc.garch.t.i)^2

    garch.t.vol[i]<-sqrt(sum(sigma.squareds.i)/10)

    garch.t.VaR[i]<--sqrt((1/df)*(df-2))*qt(p=0.99,df=df)*sqrt(sum(sigma.squareds.i)/10)

    garch.t.ES[i]<--(1/0.01)*(1/(df-1))*(df-
2+(qt(p=0.01,df=df))^2)*dt(qt(p=0.01,df=df),df=df)*sqrt(sum(sigma.squareds.i)/10)

    }*

for (i in 990:2057){

    fit.garch.t.i<-ugarchfit(spec=spec.garch.t,data=data[i:(i+250-1)],)

    w.t[i]<-coef(fit.garch.t.i)[["omega"]]

    a.t[i]<-coef(fit.garch.t.i)[["alpha"]]

    b.t[i]<-coef(fit.garch.t.i)[["beta"]]

    df<-coef(fit.garch.t.i)[["shape"]] #Kan ek hierdie df gebruik en nie weer optimering
loop nie?

    forc.garch.t.i<-ugarchforecast(fit.garch.t.i,n.ahead=10)

    sigma.squareds.i<-sigma(forc.garch.t.i)^2

    garch.t.vol[i]<-sqrt(sum(sigma.squareds.i)/10)

    garch.t.VaR[i]<--sqrt((1/df)*(df-2))*qt(p=0.99,df=df)*sqrt(sum(sigma.squareds.i)/10)

    garch.t.ES[i]<--(1/0.01)*(1/(df-1))*(df-
2+(qt(p=0.01,df=df))^2)*dt(qt(p=0.01,df=df),df=df)*sqrt(sum(sigma.squareds.i)/10)

    }

for (i in 2308:2835){

    fit.garch.t.i<-ugarchfit(spec=spec.garch.t,data=data[i:(i+250-1)],)

    w.t[i]<-coef(fit.garch.t.i)[["omega"]]

    a.t[i]<-coef(fit.garch.t.i)[["alpha"]]

    b.t[i]<-coef(fit.garch.t.i)[["beta"]]

    df<-coef(fit.garch.t.i)[["shape"]] #Kan ek hierdie df gebruik en nie weer optimering
loop nie?

    forc.garch.t.i<-ugarchforecast(fit.garch.t.i,n.ahead=10)

    sigma.squareds.i<-sigma(forc.garch.t.i)^2

    garch.t.vol[i]<-sqrt(sum(sigma.squareds.i)/10)

    garch.t.VaR[i]<--sqrt((1/df)*(df-2))*qt(p=0.99,df=df)*sqrt(sum(sigma.squareds.i)/10)

    garch.t.ES[i]<--(1/0.01)*(1/(df-1))*(df-
2+(qt(p=0.01,df=df))^2)*dt(qt(p=0.01,df=df),df=df)*sqrt(sum(sigma.squareds.i)/10)

    }

```

```

for (i in 3086:n){
  fit.garch.t.i<-ugarchfit(spec=spec.garch.t,data=data[i:(i+250-1),])
  w.t[i]<-coef(fit.garch.t.i)[["omega"]]
  a.t[i]<-coef(fit.garch.t.i)[["alpha"]]
  b.t[i]<-coef(fit.garch.t.i)[["beta1"]]
  df<-coef(fit.garch.t.i)[["shape"]] #Kan ek hierdie df gebruik en nie weer optimering
loop nie?
  forc.garch.t.i<-ugarchforecast(fit.garch.t.i,n.ahead=10)
  sigma.squareds.i<-sigma(forc.garch.t.i)^2
  garch.t.vol[i]<-sqrt(sum(sigma.squareds.i)/10)
  garch.t.VaR[i]<--sqrt((1/df)*(df-2))*qt(p=0.99,df=df)*sqrt(sum(sigma.squareds.i)/10)
  garch.t.ES[i]<--(1/0.01)*(1/(df-1))*(df-
2+(qt(p=0.01,df=df))^2)*dt(qt(p=0.01,df=df),df=df)*sqrt(sum(sigma.squareds.i)/10)
}

garch.t.vol.save<-garch.t.vol
garch.t.VaR.save<-garch.t.VaR
garch.t.ES.save<-garch.t.ES

#Multiply by square root 10 to obtain 10 day estimates:
garch.t.vol.10<-garch.t.vol.save*sqrt(10)
garch.t.VaR.10<-garch.t.VaR.save*sqrt(10)
garch.t.ES.10<-garch.t.ES.save*sqrt(10)

garch.t.vol.10.save<-garch.t.vol.10
garch.t.VaR.10.save<-garch.t.VaR.10
garch.t.ES.10.save<-garch.t.ES.10

w.t.save<-w.t
a.t.save<-a.t
b.t.save<-b.t

#Figure 8.4:
true<-data[260:4313,]

ylim<-
c(min(garch.t.vol.10.save,garch.t.VaR.10.save,garch.t.ES.10.save,true.10),max(garch.t.vol.10.s
ave,garch.t.VaR.10.save,garch.t.ES.10.save,true.10))

plot(true.10,lty=1, main="Symmetric GARCH t model",ty="l",col="light
blue",ylim=ylim,xlab="",ylab="",xaxt="n")

```

```

garch.t.vol.10[indices.empty]<-""
lines(garch.t.vol.10,lty=3)
garch.t.VaR.10[indices.empty]<-""
lines(garch.t.VaR.10,lty=2,col="red")
garch.t.ES.10[indices.empty]<-""
lines(garch.t.ES.10,lty=4,col="green")
legend(x="bottomright",legend=c("True 10 day log returns","99% 10 day VaR
forecast","Volatility forecast","99% 10 day ES"),
lty=1:4,col=c("light blue","red","black","green"))
axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.7)

par(mfrow=c(3,1))
plot(w.t.save,ty="n",main="Evolution of omega",xlab="",ylab="")
w.t[indices.empty]<-""
lines(w.t)
plot(a.t.save,ty="n",main="Evolution of alpha",xlab="",ylab="")
a.t[indices.empty]<-""
lines(a.t)
plot(b.t,ty="n",main="Evolution of beta",xlab="",ylab="")
b.t[indices.empty]<-""
lines(b.t)
par(mfrow=c(1,1))

#####
#GJR GARCH (n)
#####

spec.garch.gjr<-
ugarchspec(variance.model=list(model="gjrGARCH",garchOrder=c(1,1)),mean.model=list(armaOrder=c
(0,0)),distribution.model="norm")

garch.gjr.vol<-rep(0,n)
garch.gjr.VaR<-rep(0,n)
garch.gjr.ES<-rep(0,n)
w.gjr<-rep(0,n) #To extract the fitted parameter values
a.gjr<-rep(0,n)

```



```

b.gjr<-rep(0,n)

for (i in 1:773){

  fit.garch.gjr.i<-ugarchfit(spec=spec.garch.gjr,data=data[i:(i+250-1),])
  w.gjr[i]<-coef(fit.garch.gjr.i)[["omega"]]
  a.gjr[i]<-coef(fit.garch.gjr.i)[["alpha1"]]
  b.gjr[i]<-coef(fit.garch.gjr.i)[["beta1"]]
  forc.garch.gjr.i<-ugarchforecast(fit.garch.gjr.i,n.ahead=10)
  sigma.squareds.i<-sigma(forc.garch.gjr.i)^2
  garch.gjr.vol[i]<-sqrt(sum(sigma.squareds.i)/10)
  garch.gjr.Var[i]<-sqrt(sum(sigma.squareds.i)*(1/10))*qnorm(0.99)
  garch.gjr.ES[i]<-((1/0.01)*dnorm(qnorm(0.01))*sqrt(sum(sigma.squareds.i)*(1/10)))
}

for (i in 1024:1842){

  fit.garch.gjr.i<-ugarchfit(spec=spec.garch.gjr,data=data[i:(i+250-1),])
  w.gjr[i]<-coef(fit.garch.gjr.i)[["omega"]]
  a.gjr[i]<-coef(fit.garch.gjr.i)[["alpha1"]]
  b.gjr[i]<-coef(fit.garch.gjr.i)[["beta1"]]
  forc.garch.gjr.i<-ugarchforecast(fit.garch.gjr.i,n.ahead=10)
  sigma.squareds.i<-sigma(forc.garch.gjr.i)^2
  garch.gjr.vol[i]<-sqrt(sum(sigma.squareds.i)/10)
  garch.gjr.Var[i]<-sqrt(sum(sigma.squareds.i)*(1/10))*qnorm(0.99)
  garch.gjr.ES[i]<-((1/0.01)*dnorm(qnorm(0.01))*sqrt(sum(sigma.squareds.i)*(1/10)))
}

for (i in 2093:n){

  fit.garch.gjr.i<-ugarchfit(spec=spec.garch.gjr,data=data[i:(i+250-1),])
  w.gjr[i]<-coef(fit.garch.gjr.i)[["omega"]]
  a.gjr[i]<-coef(fit.garch.gjr.i)[["alpha1"]]
  b.gjr[i]<-coef(fit.garch.gjr.i)[["beta1"]]
  forc.garch.gjr.i<-ugarchforecast(fit.garch.gjr.i,n.ahead=10)
  sigma.squareds.i<-sigma(forc.garch.gjr.i)^2
  garch.gjr.vol[i]<-sqrt(sum(sigma.squareds.i)/10)
  garch.gjr.Var[i]<-sqrt(sum(sigma.squareds.i)*(1/10))*qnorm(0.99)
  garch.gjr.ES[i]<-((1/0.01)*dnorm(qnorm(0.01))*sqrt(sum(sigma.squareds.i)*(1/10)))
}

```

```

garch.gjr.vol.save<-garch.gjr.vol
garch.gjr.VaR.save<-garch.gjr.VaR
garch.gjr.ES.save<-garch.gjr.ES

w.gjr.save<-w.gjr
a.gjr.save<-a.gjr
b.gjr.save<-b.gjr

#Multiply by square root 10 to obtain 10 day estimates:
garch.gjr.vol.10<-garch.gjr.vol.save*sqrt(10)
garch.gjr.VaR.10<-garch.gjr.VaR.save*sqrt(10)
garch.gjr.ES.10<-garch.gjr.ES.save*sqrt(10)

garch.gjr.vol.10.save<-garch.gjr.vol.10
garch.gjr.VaR.10.save<-garch.gjr.VaR.10
garch.gjr.ES.10.save<-garch.gjr.ES.10

#Figure 8.4:
true<-data[260:4313,]

ylim<-
c(min(na.omit(garch.gjr.vol.10.save),na.omit(garch.gjr.VaR.10.save),na.omit(garch.gjr.ES.10.save,true.10)),max(na.omit(garch.gjr.vol.10.save),na.omit(garch.gjr.VaR.10.save),na.omit(garch.gjr.ES.10.save,true.10)))

plot(true.10,lty=1, main="GJR GARCH N model",ty="l",col="light
blue",ylim=ylim,xaxt="n",xlab="",ylab="")

garch.gjr.vol.10[indices.empty]<-""
lines(garch.gjr.vol.10,lty=3)

garch.gjr.VaR.10[indices.empty]<-""
lines(garch.gjr.VaR.10,lty=2,col="red")

garch.gjr.ES.10[indices.empty]<-""
lines(garch.gjr.ES.10,lty=4,col="green")

legend(x="bottomright",legend=c("True 10 day log returns","99% 10 day VaR
forecast","Volatility forecast","99% 10 day ES"),

lty=1:4,col=c("light blue","red","black","green"))

axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.8)

par(mfrow=c(3,1))

plot(w.gjr.save,ty="n",main="Evolution of omega",xlab="",ylab="")

```

```

w.gjr[indices.empty]<-" "
lines(w.gjr)
plot(a.gjr.save,ty="n",main="Evolution of alpha",xlab="",ylab="")
a.gjr[indices.empty]<-" "
lines(a.gjr)
plot(b.gjr.save,ty="n",main="Evolution of beta",xlab="",ylab="")
b.gjr[indices.empty]<-" "
lines(b.gjr)
par(mfrow=c(1,1))

#####
#GJR GARCH (t)
#####

spec.garch.gjr.t<-
ugarchspec(variance.model=list(model="gjrGARCH",garchOrder=c(1,1)),mean.model=list(armaOrder=c
(0,0)),distribution.model="std")

garch.gjr.t.vol<-rep(0,n)
garch.gjr.t.VaR<-rep(0,n)
garch.gjr.t.ES<-rep(0,n)
w.gjr.t<-rep(0,n) #To extract the fitted parameter values
a.gjr.t<-rep(0,n)
b.gjr.t<-rep(0,n)

for (i in 1:1484){
  fit.garch.gjr.t.i<-ugarchfit(spec=spec.garch.gjr.t,data=data[i:(i+250-1),])
  w.gjr.t[i]<-coef(fit.garch.gjr.t.i)[["omega"]]
  a.gjr.t[i]<-coef(fit.garch.gjr.t.i)[["alpha1"]]
  b.gjr.t[i]<-coef(fit.garch.gjr.t.i)[["beta1"]]
  df<-coef(fit.garch.gjr.t.i)[["shape"]]
  forc.garch.gjr.t.i<-ugarchforecast(fit.garch.gjr.t.i,n.ahead=10)
  sigma.squareds.i<-sigma(forc.garch.gjr.t.i)^2
  garch.gjr.t.vol[i]<-sqrt(sum(sigma.squareds.i)/10)
  garch.gjr.t.VaR[i]<-sqrt((1/df)*(df-
2))*qt(p=0.99,df=df)*sqrt(sum(sigma.squareds.i)/10)
  garch.gjr.t.ES[i]<--(1/0.01)*(1/(df-1))*(df-
2+(qt(p=0.01,df=df))^2)*dt(qt(p=0.01,df=df),df=df)*sqrt(sum(sigma.squareds.i)/10)
}

```

```

for (i in 1735:3917){
  fit.garch.gjr.t.i<-ugarchfit(spec=spec.garch.gjr.t,data=data[i:(i+250-1),])
  w.gjr.t[i]<-coef(fit.garch.gjr.t.i)[["omega"]]
  a.gjr.t[i]<-coef(fit.garch.gjr.t.i)[["alpha1"]]
  b.gjr.t[i]<-coef(fit.garch.gjr.t.i)[["beta1"]]
  df<-coef(fit.garch.gjr.t.i)[["shape"]]
  forc.garch.gjr.t.i<-ugarchforecast(fit.garch.gjr.t.i,n.ahead=10)
  sigma.squareds.i<-sigma(forc.garch.gjr.t.i)^2
  garch.gjr.t.vol[i]<-sqrt(sum(sigma.squareds.i)/10)
  garch.gjr.t.VaR[i]<-sqrt((1/df)*(df-
2))*qt(p=0.99,df=df)*sqrt(sum(sigma.squareds.i)/10)
  garch.gjr.t.ES[i]<--(1/0.01)*(1/(df-1))*(df-
2+(qt(p=0.01,df=df))^2)*dt(qt(p=0.01,df=df),df=df)*sqrt(sum(sigma.squareds.i)/10)
}

garch.gjr.t.vol.save<-garch.gjr.t.vol
garch.gjr.t.VaR.save<-garch.gjr.t.VaR
garch.gjr.t.ES.save<-garch.gjr.t.ES

w.gjr.t.save<-w.gjr.t
a.gjr.t.save<-a.gjr.t
b.gjr.t.save<-b.gjr.t

#Multiply by square root 10 to obtain 10 day estimates:
garch.gjr.t.vol.10<-garch.gjr.t.vol.save*sqrt(10)
garch.gjr.t.VaR.10<-garch.gjr.t.VaR.save*sqrt(10)
garch.gjr.t.ES.10<-garch.gjr.t.VaR.save*sqrt(10)

garch.gjr.t.vol.10.save<-garch.gjr.t.vol.10
garch.gjr.t.VaR.10.save<-garch.gjr.t.VaR.10
garch.gjr.t.ES.10.save<-garch.gjr.t.ES.10

#Figure 8.4:
true<-data[260:4313,]

ylim<-
c(min(na.omit(garch.gjr.t.vol.10.save),na.omit(garch.gjr.t.VaR.10.save),na.omit(garch.gjr.t.10
.ES.save,true.10)),max(na.omit(garch.gjr.t.vol.10.save),na.omit(garch.gjr.t.VaR.10.save),na.om
it(garch.gjr.t.ES.10.save,true.10)))

```

```

plot(true.10,lty=1, main="GJR GARCH t model",ty="l",col="light
blue",ylim=ylim,xlab="",ylab="",xaxt="n")

garch.gjr.t.vol.10[indices.empty]<-""

lines(garch.gjr.t.vol.10,lty=3)

garch.gjr.t.VaR.10[indices.empty]<-""

lines(garch.gjr.t.VaR.10,lty=2,col="red")

garch.gjr.t.ES.10[indices.empty]<-""

lines(garch.gjr.t.ES.10,lty=4,col="green")

legend(x="bottomright",legend=c("True 10 day log returns","99% 10 day VaR
forecast","Volatility forecast","99% 10 day ES"),

lty=1:4,col=c("light blue","red","black","green"))

axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.8)

par(mfrow=c(3,1))

plot(w.gjr.t.save,ty="n",main="Evolution of omega",xlab="",ylab="")

w.gjr.t[indices.empty]<-""

lines(w.gjr.t)

plot(a.gjr.t.save,ty="n",main="Evolution of alpha",xlab="",ylab="")

a.gjr.t[indices.empty]<-""

lines(a.gjr.t)

plot(b.gjr.t.save,ty="n",main="Evolution of beta",xlab="",ylab="")

b.gjr.t[indices.empty]<-""

lines(b.gjr.t)

par(mfrow=c(1,1))

#####

#EGARCH (n)

#####

spec.garch.e.n<-
ugarchspec(variance.model=list(model="eGARCH",garchOrder=c(1,1)),mean.model=list(armaOrder=c(0
,0)),distribution.model="norm")

garch.e.n.vol<-rep(0,n)

garch.e.n.VaR<-rep(0,n)

garch.e.n.ES<-rep(0,n)

w.e.n<-rep(0,n) #To extract the fitted parameter values

a.e.n<-rep(0,n)

b.e.n<-rep(0,n)

```

```

for (i in 1:3656){
  fit.garch.e.n.i<-ugarchfit(spec=spec.garch.e.n,data=data[i:(i+250-1),])
  w.e.n[i]<-coef(fit.garch.e.n.i)[["omega"]]
  a.e.n[i]<-coef(fit.garch.e.n.i)[["alpha1"]]
  b.e.n[i]<-coef(fit.garch.e.n.i)[["beta1"]]
  forc.garch.e.n.i<-ugarchforecast(fit.garch.e.n.i,n.ahead=10)
  sigma.squareds.i<-sigma(forc.garch.e.n.i)^2
  garch.e.n.vol[i]<-sqrt(sum(sigma.squareds.i)/10)
  garch.e.n.Var[i]<--sqrt(sum(sigma.squareds.i)*(1/10))*qnorm(0.99)
  garch.e.n.ES[i]<--(1/0.01)*dnorm(qnorm(0.01))*sqrt(sum(sigma.squareds.i)*(1/10))
}

for (i in 3907:n){
  fit.garch.e.n.i<-ugarchfit(spec=spec.garch.e.n,data=data[i:(i+250-1),])
  w.e.n[i]<-coef(fit.garch.e.n.i)[["omega"]]
  a.e.n[i]<-coef(fit.garch.e.n.i)[["alpha1"]]
  b.e.n[i]<-coef(fit.garch.e.n.i)[["beta1"]]
  forc.garch.e.n.i<-ugarchforecast(fit.garch.e.n.i,n.ahead=10)
  sigma.squareds.i<-sigma(forc.garch.e.n.i)^2
  garch.e.n.vol[i]<-sqrt(sum(sigma.squareds.i)/10)
  garch.e.n.Var[i]<--sqrt(sum(sigma.squareds.i)*(1/10))*qnorm(0.99)
  garch.e.n.ES[i]<--(1/0.01)*dnorm(qnorm(0.01))*sqrt(sum(sigma.squareds.i)*(1/10))
}

garch.e.n.vol.save<-garch.e.n.vol
garch.e.n.Var.save<-garch.e.n.Var
garch.e.n.ES.save<-garch.e.n.ES

w.e.n.save<-w.e.n
a.e.n.save<-a.e.n
b.e.n.save<-b.e.n

#Multiply by square root 10 to obtain 10 day estimates:
garch.e.n.vol.10<-garch.e.n.vol.save*sqrt(10)
garch.e.n.Var.10<-garch.e.n.Var.save*sqrt(10)
garch.e.n.ES.10<-garch.e.n.Var.save*sqrt(10)

```

```
garch.e.n.vol.10.save<-garch.e.n.vol.10
garch.e.n.VaR.10.save<-garch.e.n.VaR.10
garch.e.n.ES.10.save<-garch.e.n.ES.10
```

Figure 8.4:

```
true<-data[260:4313,]

ylim<-
c(min(na.omit(garch.e.n.vol.10.save),na.omit(garch.e.n.VaR.10.save),na.omit(garch.e.n.ES.10.save,true.10)),max(na.omit(garch.e.n.vol.10.save),na.omit(garch.e.n.VaR.10.save),na.omit(garch.e.n.ES.10.save,true.10)))

plot(true.10,lty=1, main="EGARCH N model",ty="l",col="light
blue",ylim=ylim,xlab="",ylab="",xaxt="n")

garch.e.n.vol.10[indices.empty]<-" "

lines(garch.e.n.vol.10,lty=3)

garch.e.n.VaR.10[indices.empty]<-" "

lines(garch.e.n.VaR.10,lty=2,col="red")

garch.e.n.ES.10[indices.empty]<-" "

lines(garch.e.n.ES.10,lty=4,col="green")

legend(x="bottomright",legend=c("True 10 day log returns","99% 10 day VaR
forecast","Volatility forecast","99% 10 day ES"),

lty=1:4,col=c("light blue","red","black","green"))

axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.8)

par(mfrow=c(3,1))

plot(w.e.n.save,ty="n",main="Evolution of omega",xlab="",ylab="")

w.e.n[indices.empty]<-" "

lines(w.e.n)

plot(a.e.n.save,ty="n",main="Evolution of alpha",xlab="",ylab="")

a.e.n[indices.empty]<-" "

lines(a.e.n)

plot(b.e.n.save,ty="n",main="Evolution of beta",xlab="",ylab="")

b.e.n[indices.empty]<-" "

lines(b.e.n)

par(mfrow=c(1,1))

#####

#EGARCH (t)
```

```

#####

spec.garch.e.t<-
ugarchspec(variance.model=list(model="eGARCH",garchOrder=c(1,1)),mean.model=list(armaOrder=c(0
,0)),distribution.model="std")

garch.e.t.vol<-rep(0,n)

garch.e.t.VaR<-rep(0,n)

garch.e.t.ES<-rep(0,n)

w.e.t<-rep(0,n) #To extract the fitted parameter values

a.e.t<-rep(0,n)

b.e.t<-rep(0,n)

for (i in 1:797){

  fit.garch.e.t.i<-ugarchfit(spec=spec.garch.e.t,data=data[i:(i+250-1)],)

  w.e.t[i]<-coef(fit.garch.e.t.i)[["omega"]]

  a.e.t[i]<-coef(fit.garch.e.t.i)[["alpha"]]

  b.e.t[i]<-coef(fit.garch.e.t.i)[["beta"]]

  df<-coef(fit.garch.e.t.i)[["shape"]] #Kan ek hierdie df gebruik en nie weer optimering
loop nie?

  forc.garch.e.t.i<-ugarchforecast(fit.garch.e.t.i,n.ahead=10)

  sigma.squareds.i<-sigma(forc.garch.e.t.i)^2

  garch.e.t.vol[i]<-sqrt(sum(sigma.squareds.i)/10)

  garch.e.t.VaR[i]<--sqrt((1/df)*(df-2))*qt(p=0.99,df=df)*sqrt(sum(sigma.squareds.i)/10)

  garch.e.t.ES[i]<--(1/0.01)*(1/(df-1))*(df-
2+(qt(p=0.01,df=df))^2)*dt(qt(p=0.01,df=df),df=df)*sqrt(sum(sigma.squareds.i)/10)

}

for (i in 1048:2740){

  fit.garch.e.t.i<-ugarchfit(spec=spec.garch.e.t,data=data[i:(i+250-1)],)

  w.e.t[i]<-coef(fit.garch.e.t.i)[["omega"]]

  a.e.t[i]<-coef(fit.garch.e.t.i)[["alpha"]]

  b.e.t[i]<-coef(fit.garch.e.t.i)[["beta"]]

  df<-coef(fit.garch.e.t.i)[["shape"]] #Kan ek hierdie df gebruik en nie weer optimering
loop nie?

  forc.garch.e.t.i<-ugarchforecast(fit.garch.e.t.i,n.ahead=10)

  sigma.squareds.i<-sigma(forc.garch.e.t.i)^2

  garch.e.t.vol[i]<-sqrt(sum(sigma.squareds.i)/10)

  garch.e.t.VaR[i]<--sqrt((1/df)*(df-2))*qt(p=0.99,df=df)*sqrt(sum(sigma.squareds.i)/10)

  garch.e.t.ES[i]<--(1/0.01)*(1/(df-1))*(df-
2+(qt(p=0.01,df=df))^2)*dt(qt(p=0.01,df=df),df=df)*sqrt(sum(sigma.squareds.i)/10)

```



```

    }

for (i in 2991:3629){

  fit.garch.e.t.i<-ugarchfit(spec=spec.garch.e.t,data=data[i:(i+250-1),])

  w.e.t[i]<-coef(fit.garch.e.t.i)[["omega"]]

  a.e.t[i]<-coef(fit.garch.e.t.i)[["alpha1"]]

  b.e.t[i]<-coef(fit.garch.e.t.i)[["beta1"]]

  df<-coef(fit.garch.e.t.i)[["shape"]] #Kan ek hierdie df gebruik en nie weer optimering
loop nie?

  forc.garch.e.t.i<-ugarchforecast(fit.garch.e.t.i,n.ahead=10)

  sigma.squareds.i<-sigma(forc.garch.e.t.i)^2

  garch.e.t.vol[i]<-sqrt(sum(sigma.squareds.i)/10)

  garch.e.t.Var[i]<--sqrt((1/df)*(df-2))*qt(p=0.99,df=df)*sqrt(sum(sigma.squareds.i)/10)

  garch.e.t.ES[i]<--(1/0.01)*(1/(df-1))*(df-
2+(qt(p=0.01,df=df))^2)*dt(qt(p=0.01,df=df),df=df)*sqrt(sum(sigma.squareds.i)/10)

  }

for (i in 3880:n){

  fit.garch.e.t.i<-ugarchfit(spec=spec.garch.e.t,data=data[i:(i+250-1),])

  w.e.t[i]<-coef(fit.garch.e.t.i)[["omega"]]

  a.e.t[i]<-coef(fit.garch.e.t.i)[["alpha1"]]

  b.e.t[i]<-coef(fit.garch.e.t.i)[["beta1"]]

  df<-coef(fit.garch.e.t.i)[["shape"]] #Kan ek hierdie df gebruik en nie weer optimering
loop nie?

  forc.garch.e.t.i<-ugarchforecast(fit.garch.e.t.i,n.ahead=10)

  sigma.squareds.i<-sigma(forc.garch.e.t.i)^2

  garch.e.t.vol[i]<-sqrt(sum(sigma.squareds.i)/10)

  garch.e.t.Var[i]<--sqrt((1/df)*(df-2))*qt(p=0.99,df=df)*sqrt(sum(sigma.squareds.i)/10)

  garch.e.t.ES[i]<--(1/0.01)*(1/(df-1))*(df-
2+(qt(p=0.01,df=df))^2)*dt(qt(p=0.01,df=df),df=df)*sqrt(sum(sigma.squareds.i)/10)

  }

garch.e.t.vol.save<-garch.e.t.vol

garch.e.t.Var.save<-garch.e.t.Var

garch.e.t.ES.save<-garch.e.t.ES

w.e.t.save<-w.e.t

a.e.t.save<-a.e.t

b.e.t.save<-b.e.t

```

```

#Multiply by square root 10 to obtain 10 day estimates:
garch.e.t.vol.10<-garch.e.t.vol.save*sqrt(10)
garch.e.t.VaR.10<-garch.e.t.VaR.save*sqrt(10)
garch.e.t.ES.10<-garch.e.t.VaR.save*sqrt(10)

garch.e.t.vol.10.save<-garch.e.t.vol.10
garch.e.t.VaR.10.save<-garch.e.t.VaR.10
garch.e.t.ES.10.save<-garch.e.t.ES.10

#Figure 8.4:
true<-data[260:4313,]

ylim<-
c(min(na.omit(garch.e.t.vol.10.save),na.omit(garch.e.t.VaR.10.save),na.omit(garch.e.t.ES.10.save,true.10)),max(na.omit(garch.e.t.vol.10.save),na.omit(garch.e.t.VaR.10.save),na.omit(garch.e.t.ES.10.save,true.10)))

plot(true.10,lty=1, main="EGARCH t model",ty="l",col="light
blue",ylim=ylim,xlab="",ylab="",xaxt="n")

garch.e.t.vol.10[indices.empty]<-""
lines(garch.e.t.vol.10,lty=3)

garch.e.t.VaR.10[indices.empty]<-""
lines(garch.e.t.VaR.10,lty=2,col="red")

garch.e.t.ES.10[indices.empty]<-""
lines(garch.e.t.ES.10,lty=4,col="green")

legend(x="bottomright",legend=c("True 10 day log returns","99% 10 day VaR
forecast","Volatility forecast","99% 10 day ES"),
lty=1:4,col=c("light blue","red","black","green"))

axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.8)

par(mfrow=c(3,1))
plot(w.e.t.save,ty="n",main="Evolution of omega",xlab="",ylab="")
w.e.t[indices.empty]<-""
lines(w.e.t)

plot(a.e.t.save,ty="n",main="Evolution of alpha",xlab="",ylab="")
a.e.t[indices.empty]<-""
lines(a.e.t)

plot(b.e.t.save,ty="n",main="Evolution of beta",xlab="",ylab="")
b.e.t[indices.empty]<-""

```

```

lines(b.e.t)

par(mfrow=c(1,1))

#####
#TO DETERMINE THE OVERLAPPING PERIODS:
#####

garch.n.vol
garch.t.vol
garch.gjr.vol
garch.gjr.t.vol
garch.e.n.vol
garch.e.t.vol

indices.work<-
(1:n)[(!(garch.n.vol==0))&(!(garch.t.vol==0))&(!(garch.gjr.vol==0))&(!(garch.gjr.t.vol==0))&(!
(garch.e.n.vol==0))&(!(garch.e.t.vol==0))]

indices.empty<-
(1:n)[(garch.n.vol==0)|(garch.t.vol==0)|(garch.gjr.vol==0)|(garch.gjr.t.vol==0)|(garch.e.n.vol
==0)|(garch.e.t.vol==0)]

#To determine the exceedences:

iid.n.VaR.x<-sum(na.omit(true.10[indices.work]<iid.n.VaR.save[indices.work]))
iid.n.ES.x<-sum(na.omit(true.10[indices.work]<iid.n.ES.save[indices.work]))
iid.t.VaR.x<-sum(na.omit(true.10[indices.work]<iid.t.VaR.save[indices.work]))
iid.t.ES.x<-sum(na.omit(true.10[indices.work]<iid.t.ES.save[indices.work]))
garch.n.VaR.x<-sum(na.omit(true.10[indices.work]<garch.n.VaR.10.save[indices.work]))
garch.n.ES.x<-sum(na.omit(true.10[indices.work]<garch.n.ES.10.save[indices.work]))
garch.t.VaR.x<-sum(na.omit(true.10[indices.work]<garch.t.VaR.10.save[indices.work]))
garch.t.ES.x<-sum(na.omit(true.10[indices.work]<garch.t.ES.10.save[indices.work]))
garch.gjr.VaR.x<-sum(na.omit(true.10[indices.work]<garch.gjr.VaR.10.save[indices.work]))
garch.gjr.ES.x<-sum(na.omit(true.10[indices.work]<garch.gjr.ES.10.save[indices.work]))
garch.gjr.t.VaR.x<-sum(na.omit(true.10[indices.work]<garch.gjr.t.VaR.10.save[indices.work]))
garch.gjr.t.ES.x<-sum(na.omit(true.10[indices.work]<garch.gjr.t.ES.10.save[indices.work]))
garch.e.n.VaR.x<-sum(na.omit(true.10[indices.work]<garch.e.n.VaR.10.save[indices.work]))
garch.e.n.ES.x<-sum(na.omit(true.10[indices.work]<garch.e.n.ES.10.save[indices.work]))
garch.e.t.VaR.x<-sum(na.omit(true.10[indices.work]<garch.e.t.VaR.10.save[indices.work]))
garch.e.t.ES.x<-sum(na.omit(true.10[indices.work]<garch.e.t.ES.10.save[indices.work]))

```

```

VaR.x<-
c(iid.n.VaR.x,iid.t.VaR.x,garch.n.VaR.x,garch.t.VaR.x,garch.gjr.VaR.x,garch.gjr.t.VaR.x,garch.
e.n.VaR.x,garch.e.t.VaR.x)

ES.x<-
c(iid.n.ES.x,iid.t.ES.x,garch.n.ES.x,garch.t.ES.x,garch.gjr.ES.x,garch.gjr.t.ES.x,garch.e.n.ES
.x,garch.e.t.ES.x)

table.x<-cbind(VaR.x,ES.x)

table.x<-rbind(c((0.01*4045),(0.01*4045)),table.x)

rownames(table.x)<-c("Expected","I.i.d. normal","I.i.d. student t","Symmetric GARCH
(n)","Symmetric GARCH (t)","GJR GARCH (n)","GJR GARCH (t)","EGARCH (n)","EGARCH (t)")

colnames(table.x)<-c("VaR","ES")

table.x

#=====
#SECTION 8.3.2.: ANALYSIS OF THE ABSOLUTE EXCEEDANCES
#=====

#VaR:

resid.iid.n.VaR<-true.10[indices.work]-iid.n.VaR.save[indices.work]
resid.iid.n.VaR<-resid.iid.n.VaR[resid.iid.n.VaR<0]
resid.iid.n.VaR<-abs(na.omit(resid.iid.n.VaR))

resid.iid.t.VaR<-true.10[indices.work]-iid.t.VaR.save[indices.work]
resid.iid.t.VaR<-resid.iid.t.VaR[resid.iid.t.VaR<0]
resid.iid.t.VaR<-abs(na.omit(resid.iid.t.VaR))

resid.garch.n.VaR<-true.10[indices.work]-garch.n.VaR.10.save[indices.work]
resid.garch.n.VaR<-resid.garch.n.VaR[resid.garch.n.VaR<0]
resid.garch.n.VaR<-abs(na.omit(resid.garch.n.VaR))

resid.garch.t.VaR<-true.10[indices.work]-garch.t.VaR.10.save[indices.work]
resid.garch.t.VaR<-resid.garch.t.VaR[resid.garch.t.VaR<0]
resid.garch.t.VaR<-abs(na.omit(resid.garch.t.VaR))

resid.garch.gjr.VaR<-true.10[indices.work]-garch.gjr.VaR.10.save[indices.work]
resid.garch.gjr.VaR<-resid.garch.gjr.VaR[resid.garch.gjr.VaR<0]
resid.garch.gjr.VaR<-abs(na.omit(resid.garch.gjr.VaR))

```

```

resid.garch.gjr.t.VaR<-true.10[indices.work]-garch.gjr.t.VaR.10.save[indices.work]
resid.garch.gjr.t.VaR<-resid.garch.gjr.t.VaR[resid.garch.gjr.t.VaR<0]
resid.garch.gjr.t.VaR<-abs(na.omit(resid.garch.gjr.t.VaR))

resid.garch.e.n.VaR<-true.10[indices.work]-garch.e.n.VaR.10.save[indices.work]
resid.garch.e.n.VaR<-resid.garch.e.n.VaR[resid.garch.e.n.VaR<0]
resid.garch.e.n.VaR<-abs(na.omit(resid.garch.e.n.VaR))

resid.garch.e.t.VaR<-true.10[indices.work]-garch.e.t.VaR.10.save[indices.work]
resid.garch.e.t.VaR<-resid.garch.e.t.VaR[resid.garch.e.t.VaR<0]
resid.garch.e.t.VaR<-abs(na.omit(resid.garch.e.t.VaR))

#Figure 8.7:
par(mfrow=c(4,2))
hist(resid.iid.n.VaR,ylab="Frequency",xlab="",main="IID N VaR")
hist(resid.iid.t.VaR,ylab="Frequency",xlab="",main="IID t VaR")
hist(resid.garch.n.VaR,ylab="Frequency",xlab="",main="Symmetric GARCH N VaR")
hist(resid.garch.t.VaR,ylab="Frequency",xlab="",main="Symmetric GARCH t VaR")
hist(resid.garch.gjr.VaR,ylab="Frequency",xlab="",main="GJR GARCH N VaR")
hist(resid.garch.gjr.t.VaR,ylab="Frequency",xlab="",main="GJR GARCH t VaR")
hist(resid.garch.e.n.VaR,ylab="Frequency",xlab="",main="EGARCH N VaR")
hist(resid.garch.e.t.VaR,ylab="Frequency",xlab="",main="EGARCH t VaR")
par(mfrow=c(1,1))

#Table 8.3:
library(tseries)
library(TSA)

means.VaR<-
c(mean(resid.iid.n.VaR),mean(resid.iid.t.VaR),mean(resid.garch.n.VaR),mean(resid.garch.t.VaR),
mean(resid.garch.gjr.VaR),

mean(resid.garch.gjr.t.VaR),mean(resid.garch.e.n.VaR),mean(resid.garch.e.t.VaR))

sdev.VaR<-
c(sd(resid.iid.n.VaR),sd(resid.iid.t.VaR),sd(resid.garch.n.VaR),sd(resid.garch.t.VaR),sd(resid
.garch.gjr.VaR),

sd(resid.garch.gjr.t.VaR),sd(resid.garch.e.n.VaR),sd(resid.garch.e.t.VaR))

medians.VaR<-
c(median(resid.iid.n.VaR),median(resid.iid.t.VaR),median(resid.garch.n.VaR),median(resid.garch
.t.VaR),median(resid.garch.gjr.VaR),

median(resid.garch.gjr.t.VaR),median(resid.garch.e.n.VaR),median(resid.garch.e.t.VaR))

```

```

max.VaR<-
c(max(resid.iid.n.VaR),max(resid.iid.t.VaR),max(resid.garch.n.VaR),max(resid.garch.t.VaR),max(
resid.garch.gjr.VaR),

max(resid.garch.gjr.t.VaR),max(resid.garch.e.n.VaR),max(resid.garch.e.t.VaR))

min.VaR<-
c(min(resid.iid.n.VaR),min(resid.iid.t.VaR),min(resid.garch.n.VaR),min(resid.garch.t.VaR),min(
resid.garch.gjr.VaR),

min(resid.garch.gjr.t.VaR),min(resid.garch.e.n.VaR),min(resid.garch.e.t.VaR))

kurt.VaR<-
c(kurtosis(resid.iid.n.VaR),kurtosis(resid.iid.t.VaR),kurtosis(resid.garch.n.VaR),kurtosis(res
id.garch.t.VaR),kurtosis(resid.garch.gjr.VaR),

kurtosis(resid.garch.gjr.t.VaR),kurtosis(resid.garch.e.n.VaR),kurtosis(resid.garch.e.t.VaR))

skew.VaR<-
c(skewness(resid.iid.n.VaR),skewness(resid.iid.t.VaR),skewness(resid.garch.n.VaR),skewness(res
id.garch.t.VaR),skewness(resid.garch.gjr.VaR),

skewness(resid.garch.gjr.t.VaR),skewness(resid.garch.e.n.VaR),skewness(resid.garch.e.t.VaR))

table.resid.VaR<-cbind(means.VaR,sdev.VaR,medians.VaR,max.VaR,min.VaR,kurt.VaR,skew.VaR)

colnames(table.resid.VaR)<-c("Mean","Std Dev","Median","Max","Min","Excess
Kurtosis","Skewness")

rownames(table.resid.VaR)<-c("IID N VaR","IID t VaR","Symmetric GARCH N VaR","Symmetric GARCH
t VaR","GJR GARCH N VaR","GJR GARCH t VaR",

"EGARCH N VaR","EGARCH t VaR")

table.resid.VaR

#ES:

resid.iid.n.ES<-true.10[indices.work]-iid.n.ES.save[indices.work]

resid.iid.n.ES<-resid.iid.n.ES[resid.iid.n.ES<0]

resid.iid.n.ES<-abs(na.omit(resid.iid.n.ES))

resid.iid.t.ES<-true.10[indices.work]-iid.t.ES.save[indices.work]

resid.iid.t.ES<-resid.iid.t.ES[resid.iid.t.ES<0]

resid.iid.t.ES<-abs(na.omit(resid.iid.t.ES))

resid.garch.n.ES<-true.10[indices.work]-garch.n.ES.10.save[indices.work]

resid.garch.n.ES<-resid.garch.n.ES[resid.garch.n.ES<0]

resid.garch.n.ES<-abs(na.omit(resid.garch.n.ES))

resid.garch.t.ES<-true.10[indices.work]-garch.t.ES.10.save[indices.work]

resid.garch.t.ES<-resid.garch.t.ES[resid.garch.t.ES<0]

resid.garch.t.ES<-abs(na.omit(resid.garch.t.ES))

```

```

resid.garch.gjr.ES<-true.10[indices.work]-garch.gjr.ES.10.save[indices.work]
resid.garch.gjr.ES<-resid.garch.gjr.ES[resid.garch.gjr.ES<0]
resid.garch.gjr.ES<-abs(na.omit(resid.garch.gjr.ES))

resid.garch.gjr.t.ES<-true.10[indices.work]-garch.gjr.t.ES.10.save[indices.work]
resid.garch.gjr.t.ES<-resid.garch.gjr.t.ES[resid.garch.gjr.t.ES<0]
resid.garch.gjr.t.ES<-abs(na.omit(resid.garch.gjr.t.ES))

resid.garch.e.n.ES<-true.10[indices.work]-garch.e.n.ES.10.save[indices.work]
resid.garch.e.n.ES<-resid.garch.e.n.ES[resid.garch.e.n.ES<0]
resid.garch.e.n.ES<-abs(na.omit(resid.garch.e.n.ES))

resid.garch.e.t.ES<-true.10[indices.work]-garch.e.t.ES.10.save[indices.work]
resid.garch.e.t.ES<-resid.garch.e.t.ES[resid.garch.e.t.ES<0]
resid.garch.e.t.ES<-abs(na.omit(resid.garch.e.t.ES))

```

#Figure 8.8:

```

par(mfrow=c(4,2))
hist(resid.iid.n.ES,ylab="Frequency",xlab="",main="IID N ES")
hist(resid.iid.t.ES,ylab="Frequency",xlab="",main="IID t ES")
hist(resid.garch.n.ES,ylab="Frequency",xlab="",main="Symmetric GARCH N ES")
hist(resid.garch.t.ES,ylab="Frequency",xlab="",main="Symmetric GARCH t ES")
hist(resid.garch.gjr.ES,ylab="Frequency",xlab="",main="GJR GARCH N ES")
hist(resid.garch.gjr.t.ES,ylab="Frequency",xlab="",main="GJR GARCH t ES")
hist(resid.garch.e.n.ES,ylab="Frequency",xlab="",main="EGARCH N ES")
hist(resid.garch.e.t.ES,ylab="Frequency",xlab="",main="EGARCH t ES")
par(mfrow=c(1,1))

```

Table 8.4:

```

library(tseries)
library(TSA)

means.ES<-
c(mean(resid.iid.n.ES),mean(resid.iid.t.ES),mean(resid.garch.n.ES),mean(resid.garch.t.ES),mean
(resid.garch.gjr.ES),

mean(resid.garch.gjr.t.ES),mean(resid.garch.e.n.ES),mean(resid.garch.e.t.ES))

sdev.ES<-
c(sd(resid.iid.n.ES),sd(resid.iid.t.ES),sd(resid.garch.n.ES),sd(resid.garch.t.ES),sd(resid.gar
ch.gjr.ES),

sd(resid.garch.gjr.t.ES),sd(resid.garch.e.n.ES),sd(resid.garch.e.t.ES))

```

```

medians.ES<-
c(median(resid.iid.n.ES),median(resid.iid.t.ES),median(resid.garch.n.ES),median(resid.garch.t.
ES),median(resid.garch.gjr.ES),

median(resid.garch.gjr.t.ES),median(resid.garch.e.n.ES),median(resid.garch.e.t.ES))

max.ES<-
c(max(resid.iid.n.ES),max(resid.iid.t.ES),max(resid.garch.n.ES),max(resid.garch.t.ES),max(resi
d.garch.gjr.ES),

max(resid.garch.gjr.t.ES),max(resid.garch.e.n.ES),max(resid.garch.e.t.ES))

min.ES<-
c(min(resid.iid.n.ES),min(resid.iid.t.ES),min(resid.garch.n.ES),min(resid.garch.t.ES),min(resi
d.garch.gjr.ES),

min(resid.garch.gjr.t.ES),min(resid.garch.e.n.ES),min(resid.garch.e.t.ES))

kurt.ES<-
c(kurtosis(resid.iid.n.ES),kurtosis(resid.iid.t.ES),kurtosis(resid.garch.n.ES),kurtosis(resid.
garch.t.ES),kurtosis(resid.garch.gjr.ES),

kurtosis(resid.garch.gjr.t.ES),kurtosis(resid.garch.e.n.ES),kurtosis(resid.garch.e.t.ES))

skew.ES<-
c(skewness(resid.iid.n.ES),skewness(resid.iid.t.ES),skewness(resid.garch.n.ES),skewness(resid.
garch.t.ES),skewness(resid.garch.gjr.ES),

skewness(resid.garch.gjr.t.ES),skewness(resid.garch.e.n.ES),skewness(resid.garch.e.t.ES))

table.resid.ES<-cbind(means.ES,sdev.ES,medians.ES,max.ES,min.ES,kurt.ES,skew.ES)

colnames(table.resid.ES)<-c("Mean","Std Dev","Median","Max","Min","Excess
Kurtosis","Skewness")

rownames(table.resid.ES)<-c("IID N ES","IID t ES","Symmetric GARCH N ES","Symmetric GARCH t
ES","GJR GARCH N ES","GJR GARCH t ES",

"EGARCH N ES","EGARCH t ES")

table.resid.ES

=====

#SECTION 8.4.: EXTREME VALUE ANALYSIS

=====

#SECTION 8.4.1.: Exploratory analysis for first data window

=====

#It may be best to use the scan() function and copy data from excel? data<-scan()

#data<-read.table(file="clipboard") #log returns from excel. I think rather use scan!

data.250.1<-scan()

data<--data.250.1 #negative log returns

dates.log.250.1<-read.table(file="clipboard")

#Figure 8.9: PLOT OF NEGATIVE LOG RETURNS:

```



```

plot(x=1:length(data),y=data,ty="l",xlab="",ylab="",main="Plot of Returns",xaxt="n")
ax.vals.4<-c(1,floor(250*spacings))
axis(side=1,at=ax.vals.4,labels=dates.log.250.1[ax.vals.4,],cex.axis=0.8)

#Figure 8.10: Histogram:
hist(data,xlab="negative log return",ylab="frequency",main="Histogram of Returns")

#Normal qqplot: Figure 8.11a
qqplot(x=rnorm(length(data)),y=data,xlab="Theoretical Quantiles",ylab="Observed Quantiles",
main="Normal QQ plot")
qqline(data)
qqplot(x=rt(n=length(data),df=5),y=data,xlab="Theoretical Quantiles",ylab="Observed
Quantiles", main="Student t QQ plot\ndf=5")
qqline(data,distribution=function(p) qt(p,df=5))
qqplot(x=rt(n=length(data),df=10),y=data,xlab="Theoretical Quantiles",ylab="Observed
Quantiles", main="Student t QQ plot\ndf=10")
qqline(data,distribution=function(p) qt(p,df=10))

#DIFFERENT ESTIMATORS OF THE EVI (Table 8.5)
#####MOMENT ESTIMATOR#####
n<-length(data)
k.vec<-c(floor(n*0.05),floor(n*0.1),floor(n*0.15),floor(n*0.2))
M.k.n<-rep(0,length(k.vec))
for (i in 1:length(k.vec)){
  gamma.hill<-mean(log(data.ordered[(n-k.vec[i]+1):n]))-log(data.ordered[(n-k.vec[i])])
  gamma.hill.2<- mean((log(data.ordered[(n-k.vec[i]+1):n]))-log(data.ordered[(n-
k.vec[i])]))^2)
  M.k.n[i]<-gamma.hill + 1 - 0.5*(1-(gamma.hill)^2/gamma.hill.2)^(-1)
}
perc<-c(0.05,0.1,0.15,0.2)
afvoer<-rbind(perc,round(k.vec,digits=0),round(M.k.n,digits=3))
rownames(afvoer)<-c("%","k","Moment estimator")
afvoer

#####GPD:MOM#####
n<-length(data)
k<-floor(length(data)*0.2)

```

```

data.GP<- data.ordered[(n-k+1):n]-data.ordered[n-k]

ybar<-mean(data.GP)

sample.var<-var(data.GP)

gamma.MOM.est<-0.5*(1-((ybar^2)/sample.var))

sigma.MOM.est<- (ybar/2)*(1+(ybar^2)/sample.var)

gamma.MOM.est

sigma.MOM.est

#estimates of s.e.'s

C.MOM<-(1-gamma.MOM.est)^2/((1-2*gamma.MOM.est)*(1-3*gamma.MOM.est)*(1-4*gamma.MOM.est))

gamma.MOM.se<- sqrt(C.MOM*(1-2*gamma.MOM.est)^2*(1-gamma.MOM.est+6*gamma.MOM.est^2))

sigma.MOM.se<- sqrt(C.MOM*2*(sigma.MOM.est^2)*(1-6*gamma.MOM.est+12*gamma.MOM.est^2))

alpha<-0.05

CI.gamma.MOM.lower<- gamma.MOM.est - qnorm(1-(alpha/2))*gamma.MOM.se

CI.gamma.MOM.upper<- gamma.MOM.est + qnorm(1-(alpha/2))*gamma.MOM.se

CI.gamma.MOM<-c(CI.gamma.MOM.lower, CI.gamma.MOM.upper)

CI.gamma.MOM

CI.sigma.MOM.lower<- sigma.MOM.est - qnorm(1-(alpha/2))*sigma.MOM.se

CI.sigma.MOM.upper<- sigma.MOM.est + qnorm(1-(alpha/2))*sigma.MOM.se

CI.sigma.MOM<-c(CI.sigma.MOM.lower, CI.sigma.MOM.upper)

CI.sigma.MOM

#####GPD:PWM#####

#PWM ESTIMATORS

k<-floor(length(data)*0.2)

data.GP<- data.ordered[(n-k+1):n]-data.ordered[n-k]

j<-1:k

gamma.PWM<-mean(((j/(k+1))^4-3)*data.GP)/mean(((j/(k+1))^2-1)*data.GP)

gamma.PWM

M100<-mean(data.GP)

M101<- mean((1-(j/(k+1)))*data.GP)

sigma.PWM<-(2*M100*M101)/(M100-2*M101)

sigma.PWM

#gamma.PWM.check<-2- (M100/(M100-2*M101))

#estimates of s.e.'s (p162)

C.PWM<-1/((1-2*gamma.PWM)*(3-2*gamma.PWM))

```

```

gamma.PWM.se<- sqrt(C.PWM*(1-gamma.PWM)*((2-gamma.PWM)^2)*(1-gamma.PWM+2*gamma.PWM^2))
sigma.PWM.se<- sqrt(C.MOM*(sigma.PWM^2)*(7-18*gamma.PWM+11*(gamma.PWM^2)-2*(gamma.PWM^3)))
alpha<-0.05

CI.gamma.PWM.lower<- gamma.PWM - qnorm(1-(alpha/2))*gamma.PWM.se
CI.gamma.PWM.upper<- gamma.PWM + qnorm(1-(alpha/2))*gamma.PWM.se
CI.gamma.PWM<-c(CI.gamma.PWM.lower, CI.gamma.PWM.upper)

CI.gamma.PWM

CI.sigma.PWM.lower<- sigma.PWM - qnorm(1-(alpha/2))*sigma.PWM.se
CI.sigma.PWM.upper<- sigma.PWM + qnorm(1-(alpha/2))*sigma.PWM.se
CI.sigma.PWM<-c(CI.sigma.PWM.lower, CI.sigma.PWM.upper)

CI.sigma.PWM

library(ismev)

#mean excess plot:
mrl.plot(data*100)

#gpd mle's:
n<-length(data)
k.vec<-c(floor(n*0.05),floor(n*0.1),floor(n*0.15),floor(n*0.2))
thresholds<-data.ordered[n-k.vec]

#The first estimates are for sigma, the second for the EVI
gpd.fit.5<-gpd.fit(data,threshold=thresholds[1],npy=250) #5%
gpd.fit.10<-gpd.fit(data,threshold=thresholds[2],npy=250) #10%
gpd.fit.15<-gpd.fit(data,threshold=thresholds[3],npy=250) #15%
gpd.fit.20<-gpd.fit(data,threshold=thresholds[4],npy=250) #20%

#diagnostics for any fitted model (specific threshold):
gpd.diag(gpd.fit.5)
gpd.diag(gpd.fit.10)
gpd.diag(gpd.fit.15)
gpd.diag(gpd.fit.20)

#profile likelihoods:
gpd.profxi(gpd.fit.20,-0.3,0.6)

locator(2)

#Line plots of the 4 different estimators (Figure 8.12):

```

```

Mkn.plot<-c(0.032,0.233,0.206,0.195)
ML.plot<-c(-0.6557,0.09459,0.09724,0.08125)
MOM.plot<-c(-0.1412,0.0541,0.0670,0.06370)
PWM.plot<-c(-0.2878,0.0762,0.06553,0.05356)

plot(x=1:4,y=Mkn.plot,ty="b",col="red",xlab="Exceedance level (%)",ylab="Estimate value",
main="EVI estimates as functions of exceedance levels",xaxt="n",
ylim=c(min(Mkn.plot,ML.plot,MOM.plot,PWM.plot),max(Mkn.plot,ML.plot,MOM.plot,PWM.plot)))
axis(side=1,at=1:4,labels=c("5%","10%","15%","20%"))
lines(ML.plot,ty="b",col="blue")
lines(MOM.plot,ty="b",col="green")
lines(PWM.plot,ty="b",col="black")
legend(x="bottomright",legend=c("Moment estimates","ML estimates","MOM estimates","PWM
estimates"),
col=c("red","blue","green","black"),lty=c(1,1,1,1))
abline(h=0,lty=2)

#Table 8.6: Hill estimators
k.vec<-c(25,37,50)
H.k.n.vec<-rep(0,length(k.vec))
for (j in 1:length(k.vec)){
  k<-k.vec[j]
  drempel<-data.ordered[250-k]
  k.largest<-data.ordered[(250-k+1):250]
  H.k.n<-mean(log(k.largest))-log(drempel)
  H.k.n.vec[j]<-H.k.n
}

#Figures 8.14-8.17: QQ plots to assess validity of GPD assumption
#Calculate the exceedances:
exceed.5<-data.ordered[(250-12+1):250]-data.ordered[250-12]
exceed.10<-data.ordered[(250-25+1):250]-data.ordered[250-25]
exceed.15<-data.ordered[(250-37+1):250]-data.ordered[250-37]
exceed.20<-data.ordered[(250-50+1):250]-data.ordered[250-50]

#(Run each part separately for each small plot)

```

```

#For the moment estimator

gamma.hat<-M.k.n[1] #for 5%

theoretical<-(1-(1:k.vec[1])/(k.vec[1]+1))^(gamma.hat)-1

par(mfrow=c(2,2))

plot(x=theoretical,y=exceed.5,main="QQ plot: Moment estimate,\n 5%
exceedances",xlab="Theoretical quantiles",ylab="Observed quantiles")

gamma.hat<-M.k.n[2] #for 10%

theoretical<-(1-(1:k.vec[2])/(k.vec[2]+1))^(gamma.hat)-1

plot(x=theoretical,y=exceed.10,main="QQ plot: Moment estimate,\n 10%
exceedances",xlab="Theoretical quantiles",ylab="Observed quantiles")

gamma.hat<-M.k.n[3] #for 15%

theoretical<-(1-(1:k.vec[3])/(k.vec[3]+1))^(gamma.hat)-1

plot(x=theoretical,y=exceed.15,main="QQ plot: Moment estimate,\n 15%
exceedances",xlab="Theoretical quantiles",ylab="Observed quantiles")

gamma.hat<-M.k.n[4] #for 20%

theoretical<-(1-(1:k.vec[4])/(k.vec[4]+1))^(gamma.hat)-1

plot(x=theoretical,y=exceed.20,main="QQ plot: Moment estimate,\n 20%
exceedances",xlab="Theoretical quantiles",ylab="Observed quantiles")

#For the ML estimator:

gamma.hat<--0.6557 #for 5%

theoretical<-(1-(1:k.vec[1])/(k.vec[1]+1))^(gamma.hat)-1

par(mfrow=c(2,2))

plot(x=theoretical,y=exceed.5,main="QQ plot: ML estimate,\n 5% exceedances",xlab="Theoretical
quantiles",ylab="Observed quantiles")

gamma.hat<-0.09459 #for 10%

theoretical<-(1-(1:k.vec[2])/(k.vec[2]+1))^(gamma.hat)-1

plot(x=theoretical,y=exceed.10,main="QQ plot: ML estimate,\n 10%
exceedances",xlab="Theoretical quantiles",ylab="Observed quantiles")

gamma.hat<-0.09724 #for 15%

theoretical<-(1-(1:k.vec[3])/(k.vec[3]+1))^(gamma.hat)-1

plot(x=theoretical,y=exceed.15,main="QQ plot: ML estimate,\n 15%
exceedances",xlab="Theoretical quantiles",ylab="Observed quantiles")

gamma.hat<-0.08125 #for 20%

theoretical<-(1-(1:k.vec[4])/(k.vec[4]+1))^(gamma.hat)-1

plot(x=theoretical,y=exceed.20,main="QQ plot: ML estimate,\n 20%
exceedances",xlab="Theoretical quantiles",ylab="Observed quantiles")

#For the MOM estimator:

gamma.hat<--0.1412 #for 5%

theoretical<-(1-(1:k.vec[1])/(k.vec[1]+1))^(gamma.hat)-1

```

```

par(mfrow=c(2,2))

plot(x=theoretical,y=exceed.5,main="QQ plot: MOM estimate,\n 5% exceedances",xlab="Theoretical
quantiles",ylab="Observed quantiles")

gamma.hat<-0.0541 #for 10%

theoretical<-(1-(1:k.vec[2])/(k.vec[2]+1))^(gamma.hat)-1

plot(x=theoretical,y=exceed.10,main="QQ plot: MOM estimate,\n 10%
exceedances",xlab="Theoretical quantiles",ylab="Observed quantiles")

gamma.hat<-0.0670 #for 15%

theoretical<-(1-(1:k.vec[3])/(k.vec[3]+1))^(gamma.hat)-1

plot(x=theoretical,y=exceed.15,main="QQ plot: MOM estimate,\n 15%
exceedances",xlab="Theoretical quantiles",ylab="Observed quantiles")

gamma.hat<-0.06370 #for 20%

theoretical<-(1-(1:k.vec[4])/(k.vec[4]+1))^(gamma.hat)-1

plot(x=theoretical,y=exceed.20,main="QQ plot: MOM estimate,\n 20%
exceedances",xlab="Theoretical quantiles",ylab="Observed quantiles")

#For the PWM estimator:

gamma.hat<--0.2878 #for 5%

theoretical<-(1-(1:k.vec[1])/(k.vec[1]+1))^(gamma.hat)-1

par(mfrow=c(2,2))

plot(x=theoretical,y=exceed.5,main="QQ plot: PWM estimate,\n 5% exceedances",xlab="Theoretical
quantiles",ylab="Observed quantiles")

gamma.hat<-0.0762 #for 10%

theoretical<-(1-(1:k.vec[2])/(k.vec[2]+1))^(gamma.hat)-1

plot(x=theoretical,y=exceed.10,main="QQ plot: PWM estimate,\n 10%
exceedances",xlab="Theoretical quantiles",ylab="Observed quantiles")

gamma.hat<-0.06553 #for 15%

theoretical<-(1-(1:k.vec[3])/(k.vec[3]+1))^(gamma.hat)-1

plot(x=theoretical,y=exceed.15,main="QQ plot: PWM estimate,\n 15%
exceedances",xlab="Theoretical quantiles",ylab="Observed quantiles")

gamma.hat<-0.05356 #for 20%

theoretical<-(1-(1:k.vec[4])/(k.vec[4]+1))^(gamma.hat)-1

plot(x=theoretical,y=exceed.20,main="QQ plot: PWM estimate,\n 20%
exceedances",xlab="Theoretical quantiles",ylab="Observed quantiles")

#=====
#SECTION 8.4.2.: VaR AND ES OVER THE WHOLE DATA SERIES
#=====

#POT approach with ML estimates

#A threshold of 20% or 50 observations will be used.

```

```

data<-J580 #This is the log returns of the Africa Financial index from 29/01/96 to 30/04/13
#But we consider the negative of the log returns in the EVT analysis:
data<--data

#First estimate parameters of GPD using ML and moving data window of 250 days.
#E.g.: Estimate for day 251 used days 1-250 of data for estimation
#Plot against time
#So we plot against day 251,252,...,4313
#Also perform the Mkn estimate for the moving window and plot

library(ismev) #adjusted gpd.fit to gpd.fit.own- give only MLEs in vector
gamma.vec<-rep(0,4063)
sigma.vec<-rep(0,4063)
M.k.n.vec<-rep(0,4063)
k<-50
for (i in 1:4063){
data.x<-data[i:(i+250-1),]
#for MLEs:
gamma.vec[i]<-gpd.fit.own(data.x,threshold=thresholds[4],npy=250)[2]
sigma.vec[i]<-gpd.fit.own(data.x,threshold=thresholds[4],npy=250)[1]
#for Mkn:
data.ordered.x<-sort(data.x,decreasing=FALSE)
gamma.hill<-mean(log(data.ordered.x[(250-k+1):250]))-log(data.ordered.x[(250-k)])
gamma.hill.2<- mean((log(data.ordered.x[(250-k+1):250])-log(data.ordered.x[(250-k)]))^2)
M.k.n.vec[i]<-gamma.hill + 1 - 0.5*(1-(gamma.hill)^2/gamma.hill.2)^(-1)
}

gamma.vec.save<-gamma.vec
sigma.vec.save<-sigma.vec
M.k.n.vec.save<-M.k.n.vec

#dates.4<-read.table(file="clipboard")
ax.vals.4<-c(1,floor(4063*spacings))

#Figure 8.18:

```

```

plot(M.k.n.vec,ty="l",ylim=c(min(M.k.n.vec,gamma.vec),max(M.k.n.vec,gamma.vec)),lty=1,col="blue",
     ,
main="Estimators of the EVI for moving data window of length 250
days",ylab="",xlab="",xaxt="n")

lines(gamma.vec,lty=2,col="red")

legend(legend=c("Moment Estimator","ML estimator of GPD
gamma"),x="topright",col=c("blue","red"),lty=1:2)

axis(side=1,at=ax.vals.4,labels=dates.4[ax.vals.4,],cex.axis=0.8)

```

Figure 8.19:

```

plot(sigma.vec,xlab="",ylab="",main="ML estimate of GPD sigma",ty="l",xaxt="n")

axis(side=1,at=ax.vals.4,labels=dates.4[ax.vals.4,],cex.axis=0.8)

```

Figure 8.20:

```

plot((sigma.vec/gamma.vec),xlab="",ylab="",main="Ratio: \nML estimate of sigma to EVI
",ty="l",xaxt="n")

axis(side=1,at=ax.vals.4,labels=dates.4[ax.vals.4,],cex.axis=0.8)

```

```

#The POT method for constructing VaR and ES (using ML estimates)

#This is for 99%

k<-50

Nu<-k

alpha<-0.01

VaR.POT.vec<-rep(0,4063) #The 1 day forecasts (250 data window)

ES.POT.vec<-rep(0,4063)

u.vec<-rep(0,4063) #Thresholds for each of the 250 day windows

for (i in 1:4063){

  data.x<-data[i:(i+250-1),]

  data.ordered.x<-sort(data.x,decreasing=FALSE)

  u<-data.ordered.x[250-k] #The threshold

  u.vec[i]<-u

  sigma<-sigma.vec[i]

  gamma<-gamma.vec[i]

  VaR.POT.vec[i]<-u+(sigma/gamma)*(((250*alpha)/Nu)^(-gamma)-1)

  ES.POT.vec[i]<-VaR.POT/(1-gamma)+(sigma-gamma*u)/(1-gamma)

}

VaR.POT.vec.save<-VaR.POT.vec

```



```

ES.POT.vec.save<-ES.POT.vec

#Now using Mkn
k<50
Nu<-k
alpha<-0.01
VaR.POT.Mkn.vec<-rep(0,4063) #The 1 day forecasts (250 data window)
ES.POT.Mkn.vec<-rep(0,4063)

for (i in 1:4063){
  data.x<-data[i:(i+250-1),]
  data.ordered.x<-sort(data.x,decreasing=FALSE)
  u<-data.ordered.x[250-k] #The threshold
  sigma<-sigma.vec[i]
  gamma<-M.k.n.vec[i]
  VaR.POT.Mkn.vec[i]<-u+(sigma/gamma)*(((250*alpha)/Nu)^(-gamma)-1)
  ES.POT.Mkn.vec[i]<-VaR.POT/(1-gamma)+(sigma-gamma*u)/(1-gamma)
}

VaR.POT.Mkn.vec.save<-VaR.POT.Mkn.vec
ES.POT.Mkn.vec.save<-ES.POT.Mkn.vec
u.vec.save<-u.vec

#Figure 8.21:
#POT and ML:
true<- J580[251:4313,]

plot(true,ty="l",col="lightblue",main="1 day POT VaR and ES estimates (ML estimate of
EVI)",xlab="",ylab="",

ylim=c(min(true,-VaR.POT.vec,-ES.POT.vec),max(true,-VaR.POT.vec,-ES.POT.vec)),xaxt="n")
lines(-VaR.POT.vec,col="red",lty=2)
lines(-ES.POT.vec,col="green",lty=3)

legend(legend=c("True daily log returns","99% 1 day VaR","99% 1 day
ES"),x="bottomright",col=c("lightblue","red","green"),lty=1:3)

axis(side=1,at=ax.vals.4,labels=dates.4[ax.vals.4,],cex.axis=0.8)

#POT and Mkn:
plot(true,ty="l",col="lightblue",main="1 day POT VaR and ES estimates (Moment estimate of
EVI)",xlab="",ylab="",

```

```

ylim=c(min(true, -VaR.POT.Mkn.vec, -ES.POT.Mkn.vec), max(true, -VaR.POT.Mkn.vec, -
ES.POT.Mkn.vec)), xaxt="n")

lines(-VaR.POT.Mkn.vec, col="red", lty=2)

lines(-ES.POT.Mkn.vec, col="green", lty=3)

legend(legend=c("True daily log returns", "99% 1 day VaR", "99% 1 day
ES"), x="bottomright", col=c("lightblue", "red", "green"), lty=1:3)

axis(side=1, at=ax.vals.4, labels=dates.4[ax.vals.4,], cex.axis=0.8)

#Table 8.7: Table of exceedances:

exceed.VaR.POT.1<-sum(true<(-VaR.POT.vec))

exceed.ES.POT.1<-sum(true<(-ES.POT.vec))

exceed.VaR.POT.Mkn.1<-sum(true<(-VaR.POT.Mkn.vec))

exceed.ES.POT.Mkn.1<-sum(true<(-ES.POT.Mkn.vec))

table.x.POT.1<-
rbind(c(40.63, 40.63), c(exceed.VaR.POT.1, exceed.ES.POT.1), c(exceed.VaR.POT.Mkn.1, exceed.ES.POT.
Mkn.1))

colnames(table.x.POT.1)<-c("VaR", "ES")

rownames(table.x.POT.1)<-c("Expected", "POT ML estimator", "POT Moment estimator")

table.x.POT.1

#SECTION 8.4.2.2. 10 DAY VaR AND ES ESTIMATES: SCALING

#To convert the one day VaR to a 10 day VaR, use the alpha root rule:

#VaR(10)=VaR(1)(10)^(gamma)

#E.g. 10 day ahead forecast made on day 251 for day 260 is...

y<-4054 #since true.10 contains 4054 returns

VaR.POT.vec.10<-VaR.POT.vec[1:y]*(10)^(gamma.vec[1:y])

VaR.POT.Mkn.vec.10<-VaR.POT.Mkn.vec[1:y]*(10)^(M.k.n.vec[1:y])

ES.POT.vec.10<-VaR.POT.vec.10/(1-gamma.vec.save[1:y])+(sigma.vec.save[1:y]-
gamma.vec.save[1:y]*u.vec.save[1:y])/(1-gamma.vec.save[1:y])

ES.POT.Mkn.vec.10<-VaR.POT.Mkn.vec.10/(1-M.k.n.vec.save[1:y])+(sigma.vec.save[1:y]-
M.k.n.vec.save[1:y]*u.vec.save[1:y])/(1-M.k.n.vec.save[1:y])

#Conversion using square root of time rule:

y<-4054 #since true.10 contains 4054 returns

VaR.POT.vec.10<-VaR.POT.vec[1:y]*(10)^(0.5)

VaR.POT.Mkn.vec.10<-VaR.POT.Mkn.vec[1:y]*(10)^(0.5)

ES.POT.vec.10<-VaR.POT.vec.10/(1-gamma.vec.save[1:y])+(sigma.vec.save[1:y]-
gamma.vec.save[1:y]*u.vec.save[1:y])/(1-gamma.vec.save[1:y])

```

```

ES.POT.Mkn.vec.10<-VaR.POT.Mkn.vec.10/(1-M.k.n.vec.save[1:y])+(sigma.vec.save[1:y]-
M.k.n.vec.save[1:y]*u.vec.save[1:y])/(1-M.k.n.vec.save[1:y])

#The empirically established scaling laws:

#First establish the historical distribution and the percentiles.

returns<-J580

hist.sd<-rep(0,4063) #The hist sd's for ALL the 250 day window periods
for (i in 1:length(hist.sd)){
hist.sd[i]<-sd(returns[i:(250+i-1),])
}

hist.quantiles.5<-quantile(hist.sd,probs=0.05)
hist.quantiles.50<-quantile(hist.sd,probs=0.5)
hist.quantiles.95<-quantile(hist.sd,probs=0.95)

hist.quantiles<-c(hist.quantiles.5,hist.quantiles.50,hist.quantiles.95)

a<-hist.quantiles[[1]]
b<-hist.quantiles[[2]]
c<-hist.quantiles[[3]]

slope1<-(0.65-0.59)/(a-b)

intercept1<-0.65-((0.65-0.59)/(a-b))*a

slope2<-(0.59-0.47)/(b-c)

intercept2<-0.59-((0.59-0.47)/(b-c))*b

#Figure 8.24: plot the function used for the interpolation:

plot(x=c(min(hist.sd),a,b,c,max(hist.sd)),y=c(0.65,0.65,0.59,0.47,0.47),ty="b",
main="Linear interpolation of scaling factors",xlab="Standard deviation",ylab="Scaling
factor")

text(x=c(a,b,c),y=c(0.65,0.59,0.47),labels=c("a","b","c"),pos=4)

legend(x="topright",legend=c("a=(0.00732,0.65)","b=(0.01213,0.59)","c=(0.02346,0.47)"))

VaR.POT.vec.10<-rep(0,4054)

ES.POT.vec.10<-rep(0,4054)

VaR.POT.Mkn.vec.10<-rep(0,4054)

ES.POT.Mkn.vec.10<-rep(0,4054)

scale.vec<-rep(0,4054)

for (i in 1:length(VaR.POT.vec.10)){

if(hist.sd[i]<a) {sc<-0.65}

if (hist.sd[i]==a) {sc<-0.65}

if((a<hist.sd[i])&(hist.sd[i]<b)) {sc<-slope1*hist.sd[i]+intercept1}

```

```

if(hist.sd[i]==b) {sc<-0.59}
if((b<hist.sd[i])&(hist.sd[i]<c)) {sc<-slope2*hist.sd[i]+intercept2}
if (hist.sd[i]==c) {sc<-0.47}
if(hist.sd[i]>=c) {sc<-0.47}

scale<-sc

scale.vec[i]<-sc

VaR.POT.vec.10.i<-VaR.POT.vec.save[i]*(10)^scale

VaR.POT.vec.10[i]<-VaR.POT.vec.10.i

VaR.POT.Mkn.vec.10.i<-VaR.POT.Mkn.vec.save[i]*(10)^scale

VaR.POT.Mkn.vec.10[i]<-VaR.POT.Mkn.vec.10.i

ES.POT.vec.10[i]<-VaR.POT.vec.10.i/(1-gamma.vec.save[i])+(sigma.vec.save[i]-
gamma.vec.save[i]*u.vec.save[i])/(1-gamma.vec.save[i])

ES.POT.Mkn.vec.10[i]<-VaR.POT.Mkn.vec.10.i/(1-M.k.n.vec.save[i])+(sigma.vec.save[i]-
M.k.n.vec.save[i]*u.vec.save[i])/(1-M.k.n.vec.save[i])

}

plot(scale.vec,cex=0.5,main="Scaling factors used",ylab="Scaling factor",xaxt="n")

axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.8)

#TEST:

i=100

if(hist.sd[i]<a) {sc<-0.65}

if (hist.sd[i]==a) {sc<-0.65}

if((a<hist.sd[i])&(hist.sd[i]<b)) {sc<-slope1*hist.sd[i]+intercept1}

if(hist.sd[i]==b) {scale<-0.59}

if((b<hist.sd[i])&(hist.sd[i]<c)) {sc<-slope2*hist.sd[i]+intercept2}

if (hist.sd[i]==c) {sc<-0.47}

if(hist.sd[i]>=c) {sc<-0.47}

scale<-sc

scale

#Saving the values of ES and VaR for each of the 3 methods (and the ML and Mkn estimates):

#Square root of time rule:

VaR.POT.vec.10.SquareRoot<-VaR.POT.vec.10

ES.POT.vec.10.SquareRoot<-ES.POT.vec.10

VaR.POT.Mkn.vec.10.SquareRoot<-VaR.POT.Mkn.vec.10

ES.POT.Mkn.vec.10.SquareRoot<-ES.POT.Mkn.vec.10

#Alpha root rule:

```

```

VaR.POT.vec.10.AlphaRoot<-VaR.POT.vec.10
ES.POT.vec.10.AlphaRoot<-ES.POT.vec.10
VaR.POT.Mkn.vec.10.AlphaRoot<-VaR.POT.Mkn.vec.10
ES.POT.Mkn.vec.10.AlphaRoot<-ES.POT.Mkn.vec.10

#Empirically established scaling laws (McNeil):
VaR.POT.vec.10.McNeil<-VaR.POT.vec.10
ES.POT.vec.10.McNeil<-ES.POT.vec.10
VaR.POT.Mkn.vec.10.McNeil<-VaR.POT.Mkn.vec.10
ES.POT.Mkn.vec.10.McNeil<-ES.POT.Mkn.vec.10

#Figure 8.23:
#THE PLOTS: (Re-run for each method!)
#For the ML estimate of the EVI:
plot(true.10,ty="l",col="lightblue",main="10 day POT VaR and ES estimates (ML estimate of
EVI)",xlab="",ylab="",
ylim=c(min(true.10,-VaR.POT.vec.10,-ES.POT.vec.10),max(true.10,-VaR.POT.vec.10,-
ES.POT.vec.10)),xaxt="n")
lines(-VaR.POT.vec.10,col="red",lty=2)
lines(-ES.POT.vec.10,col="green",lty=3)
legend(legend=c("True 10 day returns","99% 10 day VaR","99% 10 day
ES"),x="topright",col=c("lightblue","red","green"),lty=1:3)
axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.8)

#For the Mkn estimate of the EVI:
plot(true.10,ty="l",col="lightblue",main="10 day POT VaR and ES estimates (Moment of
EVI)",xlab="",ylab="",
ylim=c(min(true.10,-VaR.POT.vec.10,-VaR.POT.Mkn.vec.10,-ES.POT.Mkn.vec.10),max(true.10,-
VaR.POT.Mkn.vec.10,-ES.POT.Mkn.vec.10)),xaxt="n")
lines(-VaR.POT.Mkn.vec.10,col="red",lty=2)
lines(-ES.POT.Mkn.vec.10,col="green",lty=3)
legend(legend=c("True 10 day returns","99% 10 day VaR","99% 10 day
ES"),x="topright",col=c("lightblue","red","green"),lty=1:3)
axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.8)

#Figure 8.27:
#Comparative plots on same axis:
#VaR+ML for 3 different methods:
plot(true.10,ty="l",col="lightblue",main="99% 10 day POT VaR estimates (ML estimate of
EVI)",xlab="",ylab="",

```

```

ylim=c(min(true.10,-VaR.POT.vec.10.SquareRoot,-VaR.POT.vec.10.AlphaRoot,-
VaR.POT.vec.10.McNeil),

max(true.10,-VaR.POT.vec.10.SquareRoot,-VaR.POT.vec.10.AlphaRoot,-
VaR.POT.vec.10.McNeil)),xaxt="n")

lines(-VaR.POT.vec.10.SquareRoot,col="red",lty=2)

lines(-VaR.POT.vec.10.AlphaRoot,col="darkgreen",lty=3)

lines(-VaR.POT.vec.10.McNeil,col="purple",lty=4)

legend(legend=c("True 10 day log returns","Square Root","Alpha Root","Empirical scaling
laws"),x="bottomright",col=c("lightblue","red","darkgreen","purple"),lty=1:4)

axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.8)

#VaR+Mkn for 3 different methods:

plot(true.10,ty="l",col="lightblue",main="99% 10 day POT VaR estimates (Moment estimate of
EVI)",xlab="",ylab="",

ylim=c(min(true.10,-VaR.POT.Mkn.vec.10.SquareRoot,-VaR.POT.Mkn.vec.10.AlphaRoot,-
VaR.POT.Mkn.vec.10.McNeil),

max(true.10,-VaR.POT.Mkn.vec.10.SquareRoot,-VaR.POT.Mkn.vec.10.AlphaRoot,-
VaR.POT.Mkn.vec.10.McNeil)),xaxt="n")

lines(-VaR.POT.Mkn.vec.10.SquareRoot,col="red",lty=2)

lines(-VaR.POT.Mkn.vec.10.AlphaRoot,col="darkgreen",lty=3)

lines(-VaR.POT.Mkn.vec.10.McNeil,col="purple",lty=4)

legend(legend=c("True 10 day log returns","Square Root","Alpha Root","Empirical scaling
laws"),x="bottomright",col=c("lightblue","red","darkgreen","purple"),lty=1:4)

axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.8)

#ES+ML for 3 different methods:

plot(true.10,ty="l",col="lightblue",main="99% 10 day POT ES estimates (ML estimate of
EVI)",xlab="",ylab="",

ylim=c(min(true.10,-ES.POT.vec.10.SquareRoot,-ES.POT.vec.10.AlphaRoot,-ES.POT.vec.10.McNeil),

max(true.10,-ES.POT.vec.10.SquareRoot,-ES.POT.vec.10.AlphaRoot,-
ES.POT.vec.10.McNeil)),xaxt="n")

lines(-ES.POT.vec.10.SquareRoot,col="red",lty=2)

lines(-ES.POT.vec.10.AlphaRoot,col="darkgreen",lty=3)

lines(-ES.POT.vec.10.McNeil,col="purple",lty=4)

legend(legend=c("True 10 day log returns","Square Root","Alpha Root","Empirical scaling
laws"),x="bottomright",col=c("lightblue","red","darkgreen","purple"),lty=1:4)

axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.8)

#ES+Mkn for 3 different methods:

```

```

plot(true.10,ty="l",col="lightblue",main="99% 10 day POT ES estimates (Moment estimate of
EVI)",xlab="",ylab="",

ylim=c(min(true.10,-ES.POT.Mkn.vec.10.SquareRoot,-ES.POT.Mkn.vec.10.AlphaRoot,-
ES.POT.Mkn.vec.10.McNeil),

max(true.10,-ES.POT.Mkn.vec.10.SquareRoot,-ES.POT.Mkn.vec.10.AlphaRoot,-
ES.POT.Mkn.vec.10.McNeil)),xaxt="n")

lines(-ES.POT.Mkn.vec.10.SquareRoot,col="red",lty=2)

lines(-ES.POT.Mkn.vec.10.AlphaRoot,col="darkgreen",lty=3)

lines(-ES.POT.Mkn.vec.10.McNeil,col="purple",lty=4)

legend(legend=c("True 10 day log returns","Square Root","Alpha Root","Empirical scaling
laws"),x="bottomright",col=c("lightblue","red","darkgreen","purple"),lty=1:4)

axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.8)

#Table 8.7:
violations.VaR.POT<-sum(true.10<(-VaR.POT.vec.10))
violations.ES.POT<-sum(true.10<(-ES.POT.vec.10))
#violations.VaR.POT<-sum(true.10<(-VaR.POT.vec.10))
#violations.ES.POT<-sum(true.10<(-ES.POT.vec.10))

table<-rbind(40.45,violations.VaR.POT,violations.ES.POT)
colnames(table)<-"Number of violations"
rownames(table)<-c("Expected","VaR","ES")
table

violations.VaR.POT.Mkn<-sum(true.10<(-VaR.POT.Mkn.vec.10[1:length(true.10)]))
violations.ES.POT.Mkn<-sum(true.10<(-ES.POT.Mkn.vec.10[1:length(true.10)]))
#violations.VaR.POT.Mkn<-sum(true.10<(-VaR.POT.Mkn.vec.10))
#violations.ES.POT.Mkn<-sum(true.10<(-ES.POT.Mkn.vec.10))

table<-rbind(40.45,violations.VaR.POT.Mkn,violations.ES.POT.Mkn)
colnames(table)<-"Number of violations"
rownames(table)<-c("Expected","VaR","ES")
table

#=====
#SECTION 8.4.3: Augmenting GARCH forecasts with EVT forecasts
#=====

```

```

#Plots (Figures 8.28-8.30):

#####SYMMETRIC GARCH#####

#4 plots: N with ML, N with Mkn, t with ML, t with Mkn

par(mfrow=c(3,2))

plot(1:10,1:10,ty="n",xaxt="n",yaxt="n",xlab="",ylab="",main="SYMMETRIC GARCH")

legend(x="topleft",legend=c("True 10 day returns","99% 10 day VaR(GARCH)","99% 10 day ES
(GARCH)","99% 10 day VaR (POT)","99% 10 day ES (POT)"),

lty=c(1,2,4,2,4),lwd=c(1,1,1,2,2),col=c("light blue","red","green","darkred","darkgreen"))

#SYMM N WITH ML

#The GARCH part:

ylim<-c(min(garch.n.VaR.10.save,garch.n.ES.10.save,-VaR.POT.vec.10.McNeil,-
ES.POT.vec.10.McNeil,true.10),

max(garch.n.VaR.10.save,garch.n.ES.10.save,-VaR.POT.vec.10.McNeil, -
ES.POT.vec.10.McNeil,true.10))

plot(true.10,lty=1, main="Forecasts from symmetric GARCH N model\ncombined with POT (ML
estimate of EVI)",ty="l",col="light blue",ylim=ylim,xaxt="n",xlab="",ylab="")

garch.n.VaR.10.add<-garch.n.VaR.10.save
garch.n.ES.10.add<-garch.n.ES.10.save
garch.n.VaR.10.add[indices.empty]<-""
lines(garch.n.VaR.10.add,lty=2,col="red")
garch.n.ES.10.add[indices.empty]<-""
lines(garch.n.ES.10.add,lty=4,col="green")

#legend(x="bottomright",legend=c("True returns","99% 10 day VaR(GARCH)","99% 10 day ES
(GARCH)","99% 10 day VaR (POT)","99% 10 day ES (POT)"),

#lty=c(1,2,4,2,4),lwd=c(1,1,1,2,2),col=c("light blue","red","green","darkred","darkgreen"))

#The EVT part:

VaR.POT.vec.10.McNeil.add<--VaR.POT.vec.10.McNeil
ES.POT.vec.10.McNeil.add<--ES.POT.vec.10.McNeil
VaR.POT.vec.10.McNeil.add[indices.work]<-""
ES.POT.vec.10.McNeil.add[indices.work]<-""
lines(VaR.POT.vec.10.McNeil.add,lty=2,lwd=2,col="darkred")
lines(ES.POT.vec.10.McNeil.add,lty=4,lwd=2,col="darkgreen")

axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.8)

#SYMM N WITH Mkn

```



```

#The GARCH part:

ylim<-c(min(garch.n.VaR.10.save,garch.n.ES.10.save,-VaR.POT.Mkn.vec.10.McNeil,-
ES.POT.Mkn.vec.10.McNeil,true.10),

max(garch.n.VaR.10.save,garch.n.ES.10.save,-VaR.POT.Mkn.vec.10.McNeil, -
ES.POT.Mkn.vec.10.McNeil,true.10))

plot(true.10,lty=1, main="Forecasts from symmetric GARCH N model\ncombined with POT (Moment
estimate of EVI)",ty="l",col="light blue",ylim=ylim,xaxt="n",xlab="",ylab="")

garch.n.VaR.10.add<-garch.n.VaR.10.save

garch.n.ES.10.add<-garch.n.ES.10.save

garch.n.VaR.10.add[indices.empty]<-""

lines(garch.n.VaR.10.add,lty=2,col="red")

garch.n.ES.10.add[indices.empty]<-""

lines(garch.n.ES.10.add,lty=4,col="green")

#legend(x="bottomright",legend=c("True returns","99% 10 day VaR(GARCH)","99% 10 day ES
(GARCH)","99% 10 day VaR (POT)","99% 10 day ES (POT)"),

#lty=c(1,2,4,2,4),lwd=c(1,1,1,2,2),col=c("light blue","red","green","darkred","darkgreen"))

#The EVT part:

VaR.POT.Mkn.vec.10.McNeil.add<--VaR.POT.Mkn.vec.10.McNeil

ES.POT.Mkn.vec.10.McNeil.add<--ES.POT.Mkn.vec.10.McNeil

VaR.POT.Mkn.vec.10.McNeil.add[indices.work]<-""

ES.POT.Mkn.vec.10.McNeil.add[indices.work]<-""

lines(VaR.POT.Mkn.vec.10.McNeil.add,lty=2,lwd=2,col="darkred")

lines(ES.POT.Mkn.vec.10.McNeil.add,lty=4,lwd=2,col="darkgreen")

axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.8)

#SYMM t WITH ML

#The GARCH part:

ylim<-c(min(garch.t.VaR.10.save,garch.t.ES.10.save,-VaR.POT.vec.10.McNeil,-
ES.POT.vec.10.McNeil,true.10),

max(garch.t.VaR.10.save,garch.t.ES.10.save,-VaR.POT.vec.10.McNeil, -
ES.POT.vec.10.McNeil,true.10))

plot(true.10,lty=1, main="Forecasts from symmetric GARCH Student t model\ncombined with POT
(ML estimate of EVI)",ty="l",col="light blue",ylim=ylim,xaxt="n",xlab="",ylab="")

garch.t.VaR.10.add<-garch.t.VaR.10.save

garch.t.ES.10.add<-garch.t.ES.10.save

garch.t.VaR.10.add[indices.empty]<-""

lines(garch.t.VaR.10.add,lty=2,col="red")

garch.t.ES.10.add[indices.empty]<-""

```

```

lines(garch.t.ES.10.add,lty=4,col="green")

#legend(x="bottomright",legend=c("True returns","99% 10 day VaR(GARCH)","99% 10 day ES
(GARCH)","99% 10 day VaR (POT)","99% 10 day ES (POT)"),

#lty=c(1,2,4,2,4),lwd=c(1,1,1,2,2),col=c("light blue","red","green","darkred","darkgreen"))

#The EVT part:

VaR.POT.vec.10.McNeil.add<--VaR.POT.vec.10.McNeil

ES.POT.vec.10.McNeil.add<--ES.POT.vec.10.McNeil

VaR.POT.vec.10.McNeil.add[indices.work]<-" "

ES.POT.vec.10.McNeil.add[indices.work]<-" "

lines(VaR.POT.vec.10.McNeil.add,lty=2,lwd=2,col="darkred")

lines(ES.POT.vec.10.McNeil.add,lty=4,lwd=2,col="darkgreen")

axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.8)

#SYMM t WITH Mkn

#The GARCH part:

ylim<-c(min(garch.t.VaR.10.save,garch.t.ES.10.save,-VaR.POT.Mkn.vec.10.McNeil,-
ES.POT.Mkn.vec.10.McNeil,true.10),

max(garch.t.VaR.10.save,garch.t.ES.10.save,-VaR.POT.Mkn.vec.10.McNeil,-
ES.POT.Mkn.vec.10.McNeil,true.10))

plot(true.10,lty=1,main="Forecasts from symmetric GARCH Student t model\ncombined with POT
(Moment estimate of EVI)",ty="l",col="light blue",ylim=ylim,xaxt="n",xlab="",ylab="")

garch.t.VaR.10.add<-garch.t.VaR.10.save

garch.t.ES.10.add<-garch.t.ES.10.save

garch.t.VaR.10.add[indices.empty]<-" "

lines(garch.t.VaR.10.add,lty=2,col="red")

garch.t.ES.10.add[indices.empty]<-" "

lines(garch.t.ES.10.add,lty=4,col="green")

#legend(x="bottomright",legend=c("True returns","99% 10 day VaR(GARCH)","99% 10 day ES
(GARCH)","99% 10 day VaR (POT)","99% 10 day ES (POT)"),

#lty=c(1,2,4,2,4),lwd=c(1,1,1,2,2),col=c("light blue","red","green","darkred","darkgreen"))

#The EVT part:

VaR.POT.Mkn.vec.10.McNeil.add<--VaR.POT.Mkn.vec.10.McNeil

ES.POT.Mkn.vec.10.McNeil.add<--ES.POT.Mkn.vec.10.McNeil

VaR.POT.Mkn.vec.10.McNeil.add[indices.work]<-" "

ES.POT.Mkn.vec.10.McNeil.add[indices.work]<-" "

lines(VaR.POT.Mkn.vec.10.McNeil.add,lty=2,lwd=2,col="darkred")

lines(ES.POT.Mkn.vec.10.McNeil.add,lty=4,lwd=2,col="darkgreen")

axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.8)

```

```

#####GJR GARCH#####

#4 plots: N with ML, N with Mkn, t with ML, t with Mkn

par(mfrow=c(3,2))

plot(1:10,1:10,ty="n",xaxt="n",yaxt="n",xlab="",ylab="",main="GJR GARCH")

legend(x="topleft",legend=c("True 10 day returns","99% 10 day VaR(GARCH)","99% 10 day ES
(GARCH)","99% 10 day VaR (POT)","99% 10 day ES (POT)"),

lty=c(1,2,4,2,4),lwd=c(1,1,1,2,2),col=c("light blue","red","green","darkred","darkgreen"))

#GJR GARCH N WITH ML

#The GARCH part:

ylim<-c(min(na.omit(garch.gjr.VaR.10.save),na.omit(garch.gjr.ES.10.save),-
VaR.POT.vec.10.McNeil,-ES.POT.vec.10.McNeil,true.10),

max(na.omit(garch.gjr.VaR.10.save),na.omit(garch.gjr.ES.10.save),-VaR.POT.vec.10.McNeil, -
ES.POT.vec.10.McNeil,true.10))

plot(true.10,lty=1, main="Forecasts from GJR GARCH N model\ncombined with POT (ML estimate of
EVI)",ty="l",col="light blue",ylim=ylim,xaxt="n",xlab="",ylab="")

garch.gjr.VaR.10.add<-garch.gjr.VaR.10.save

garch.gjr.ES.10.add<-garch.gjr.ES.10.save

garch.gjr.VaR.10.add[indices.empty]<-" "

lines(garch.gjr.VaR.10.add,lty=2,col="red")

garch.gjr.ES.10.add[indices.empty]<-" "

lines(garch.gjr.ES.10.add,lty=4,col="green")

#legend(x="bottomright",legend=c("True returns","99% 10 day VaR(GARCH)","99% 10 day ES
(GARCH)","99% 10 day VaR (POT)","99% 10 day ES (POT)"),

#lty=c(1,2,4,2,4),lwd=c(1,1,1,2,2),col=c("light blue","red","green","darkred","darkgreen"))

#The EVT part:

VaR.POT.vec.10.McNeil.add<--VaR.POT.vec.10.McNeil

ES.POT.vec.10.McNeil.add<--ES.POT.vec.10.McNeil

VaR.POT.vec.10.McNeil.add[indices.work]<-" "

ES.POT.vec.10.McNeil.add[indices.work]<-" "

lines(VaR.POT.vec.10.McNeil.add,lty=2,lwd=2,col="darkred")

lines(ES.POT.vec.10.McNeil.add,lty=4,lwd=2,col="darkgreen")

axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.8)

#GJR GARCH N WITH Mkn

```

```

#The GARCH part:

ylim<-c(min(na.omit(garch.gjr.VaR.10.save),na.omit(garch.gjr.ES.10.save),-
VaR.POT.Mkn.vec.10.McNeil,-ES.POT.Mkn.vec.10.McNeil,true.10),

max(na.omit(garch.gjr.VaR.10.save),na.omit(garch.gjr.ES.10.save),-VaR.POT.Mkn.vec.10.McNeil, -
ES.POT.Mkn.vec.10.McNeil,true.10))

plot(true.10,lty=1, main="Forecasts from GJR GARCH N model\ncombined with POT (Moment estimate
of EVI)",ty="l",col="light blue",ylim=ylim,xaxt="n",xlab="",ylab="")

garch.gjr.VaR.10.add<-garch.gjr.VaR.10.save

garch.gjr.ES.10.add<-garch.gjr.ES.10.save

garch.gjr.VaR.10.add[indices.empty]<-" "

lines(garch.gjr.VaR.10.add,lty=2,col="red")

garch.gjr.ES.10.add[indices.empty]<-" "

lines(garch.gjr.ES.10.add,lty=4,col="green")

#legend(x="bottomright",legend=c("True returns","99% 10 day VaR(GARCH)","99% 10 day ES
(GARCH)","99% 10 day VaR (POT)","99% 10 day ES (POT)"),

#lty=c(1,2,4,2,4),lwd=c(1,1,1,2,2),col=c("light blue","red","green","darkred","darkgreen"))

#The EVT part:

VaR.POT.Mkn.vec.10.McNeil.add<--VaR.POT.Mkn.vec.10.McNeil

ES.POT.Mkn.vec.10.McNeil.add<--ES.POT.Mkn.vec.10.McNeil

VaR.POT.Mkn.vec.10.McNeil.add[indices.work]<-" "

ES.POT.Mkn.vec.10.McNeil.add[indices.work]<-" "

lines(VaR.POT.Mkn.vec.10.McNeil.add,lty=2,lwd=2,col="darkred")

lines(ES.POT.Mkn.vec.10.McNeil.add,lty=4,lwd=2,col="darkgreen")

axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.8)

#GJR GARCH t WITH ML

#The GARCH part:

ylim<-c(min(na.omit(garch.gjr.t.VaR.10.save),na.omit(garch.gjr.t.ES.10.save),-
VaR.POT.vec.10.McNeil,-ES.POT.vec.10.McNeil,true.10),

max(na.omit(garch.gjr.t.VaR.10.save),na.omit(garch.gjr.t.ES.10.save),-VaR.POT.vec.10.McNeil, -
ES.POT.vec.10.McNeil,true.10))

plot(true.10,lty=1, main="Forecasts from GJR GARCH Student t model\ncombined with POT (ML
estimate of EVI)",ty="l",col="light blue",ylim=ylim,xaxt="n",xlab="",ylab="")

garch.gjr.t.VaR.10.add<-garch.gjr.t.VaR.10.save

garch.gjr.t.ES.10.add<-garch.gjr.t.ES.10.save

garch.gjr.t.VaR.10.add[indices.empty]<-" "

lines(garch.gjr.t.VaR.10.add,lty=2,col="red")

garch.gjr.t.ES.10.add[indices.empty]<-" "

lines(garch.gjr.t.ES.10.add,lty=4,col="green")

```

```

#legend(x="bottomright",legend=c("True returns","99% 10 day VaR(GARCH)","99% 10 day ES
(GARCH)","99% 10 day VaR (POT)","99% 10 day ES (POT)"),

#lty=c(1,2,4,2,4),lwd=c(1,1,1,2,2),col=c("light blue","red","green","darkred","darkgreen"))

#The EVT part:

VaR.POT.vec.10.McNeil.add<--VaR.POT.vec.10.McNeil

ES.POT.vec.10.McNeil.add<--ES.POT.vec.10.McNeil

VaR.POT.vec.10.McNeil.add[indices.work]<-" "

ES.POT.vec.10.McNeil.add[indices.work]<-" "

lines(VaR.POT.vec.10.McNeil.add,lty=2,lwd=2,col="darkred")

lines(ES.POT.vec.10.McNeil.add,lty=4,lwd=2,col="darkgreen")

axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.8)

#GJR GARCH t WITH Mkn

#The GARCH part:

ylim<-c(min(na.omit(garch.gjr.t.VaR.10.save),na.omit(garch.gjr.t.ES.10.save),-
VaR.POT.Mkn.vec.10.McNeil,-ES.POT.Mkn.vec.10.McNeil,true.10),

max(na.omit(garch.gjr.t.VaR.10.save),na.omit(garch.gjr.t.ES.10.save),-
VaR.POT.Mkn.vec.10.McNeil,-ES.POT.Mkn.vec.10.McNeil,true.10))

plot(true.10,lty=1,main="Forecasts from GJR GARCH Student t model\ncombined with POT (Moment
estimate of EVI)",ty="l",col="light blue",ylim=ylim,xaxt="n",xlab="",ylab="")

garch.gjr.t.VaR.10.add<-garch.gjr.t.VaR.10.save

garch.gjr.t.ES.10.add<-garch.gjr.t.ES.10.save

garch.gjr.t.VaR.10.add[indices.empty]<-" "

lines(garch.gjr.t.VaR.10.add,lty=2,col="red")

garch.gjr.t.ES.10.add[indices.empty]<-" "

lines(garch.gjr.t.ES.10.add,lty=4,col="green")

#legend(x="bottomright",legend=c("True returns","99% 10 day VaR(GARCH)","99% 10 day ES
(GARCH)","99% 10 day VaR (POT)","99% 10 day ES (POT)"),

#lty=c(1,2,4,2,4),lwd=c(1,1,1,2,2),col=c("light blue","red","green","darkred","darkgreen"))

#The EVT part:

VaR.POT.Mkn.vec.10.McNeil.add<--VaR.POT.Mkn.vec.10.McNeil

ES.POT.Mkn.vec.10.McNeil.add<--ES.POT.Mkn.vec.10.McNeil

VaR.POT.Mkn.vec.10.McNeil.add[indices.work]<-" "

ES.POT.Mkn.vec.10.McNeil.add[indices.work]<-" "

lines(VaR.POT.Mkn.vec.10.McNeil.add,lty=2,lwd=2,col="darkred")

lines(ES.POT.Mkn.vec.10.McNeil.add,lty=4,lwd=2,col="darkgreen")

axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.8)

```

```
#####EGARCH#####

#4 plots: N with ML, N with Mkn, t with ML, t with Mkn

par(mfrow=c(3,2))

plot(1:10,1:10,ty="n",xaxt="n",yaxt="n",xlab="",ylab="",main="EGARCH")

legend(x="topleft",legend=c("True 10 day returns","99% 10 day VaR(GARCH)","99% 10 day ES
(GARCH)","99% 10 day VaR (POT)","99% 10 day ES (POT)"),

lty=c(1,2,4,2,4),lwd=c(1,1,1,2,2),col=c("light blue","red","green","darkred","darkgreen"))

#EGARCH N WITH ML

#The GARCH part:

ylim<-c(min(na.omit(garch.e.n.VaR.10.save),na.omit(garch.e.n.ES.10.save),-
VaR.POT.vec.10.McNeil,-ES.POT.vec.10.McNeil,true.10),

max(na.omit(garch.e.n.VaR.10.save),na.omit(garch.e.n.ES.10.save),-VaR.POT.vec.10.McNeil, -
ES.POT.vec.10.McNeil,true.10))

plot(true.10,lty=1, main="Forecasts from EGARCH N model\ncombined with POT (ML estimate of
EVI)",ty="l",col="light blue",ylim=ylim,xaxt="n",xlab="",ylab="")

garch.e.n.VaR.10.add<-garch.e.n.VaR.10.save

garch.e.n.ES.10.add<-garch.e.n.ES.10.save

garch.e.n.VaR.10.add[indices.empty]<-""

lines(garch.e.n.VaR.10.add,lty=2,col="red")

garch.e.n.ES.10.add[indices.empty]<-""

lines(garch.e.n.ES.10.add,lty=4,col="green")

#legend(x="bottomright",legend=c("True returns","99% 10 day VaR(GARCH)","99% 10 day ES
(GARCH)","99% 10 day VaR (POT)","99% 10 day ES (POT)"),

#lty=c(1,2,4,2,4),lwd=c(1,1,1,2,2),col=c("light blue","red","green","darkred","darkgreen"))

#The EVT part:

VaR.POT.vec.10.McNeil.add<--VaR.POT.vec.10.McNeil

ES.POT.vec.10.McNeil.add<--ES.POT.vec.10.McNeil

VaR.POT.vec.10.McNeil.add[indices.work]<-""

ES.POT.vec.10.McNeil.add[indices.work]<-""

lines(VaR.POT.vec.10.McNeil.add,lty=2,lwd=2,col="darkred")

lines(ES.POT.vec.10.McNeil.add,lty=4,lwd=2,col="darkgreen")

axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.8)
```

```

#EGARCH N WITH Mkn

#The GARCH part:

ylim<-c(min(na.omit(garch.e.n.VaR.10.save),na.omit(garch.e.n.ES.10.save),-
VaR.POT.Mkn.vec.10.McNeil,-ES.POT.Mkn.vec.10.McNeil,true.10),

max(na.omit(garch.e.n.VaR.10.save),na.omit(garch.e.n.ES.10.save),-VaR.POT.Mkn.vec.10.McNeil, -
ES.POT.Mkn.vec.10.McNeil,true.10))

plot(true.10,lty=1, main="Forecasts from EGARCH N model\ncombined with POT (Moment estimate of
EVI)",ty="l",col="light blue",ylim=ylim,xaxt="n",xlab="",ylab="")

garch.e.n.VaR.10.add<-garch.e.n.VaR.10.save

garch.e.n.ES.10.add<-garch.e.n.ES.10.save

garch.e.n.VaR.10.add[indices.empty]<-" "

lines(garch.e.n.VaR.10.add,lty=2,col="red")

garch.e.n.ES.10.add[indices.empty]<-" "

lines(garch.e.n.ES.10.add,lty=4,col="green")

#legend(x="bottomright",legend=c("True returns","99% 10 day VaR(GARCH)","99% 10 day ES
(GARCH)","99% 10 day VaR (POT)","99% 10 day ES (POT)"),

#lty=c(1,2,4,2,4),lwd=c(1,1,1,2,2),col=c("light blue","red","green","darkred","darkgreen"))

#The EVT part:

VaR.POT.Mkn.vec.10.McNeil.add<--VaR.POT.Mkn.vec.10.McNeil

ES.POT.Mkn.vec.10.McNeil.add<--ES.POT.Mkn.vec.10.McNeil

VaR.POT.Mkn.vec.10.McNeil.add[indices.work]<-" "

ES.POT.Mkn.vec.10.McNeil.add[indices.work]<-" "

lines(VaR.POT.Mkn.vec.10.McNeil.add,lty=2,lwd=2,col="darkred")

lines(ES.POT.Mkn.vec.10.McNeil.add,lty=4,lwd=2,col="darkgreen")

axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.8)

#EGARCH t WITH ML

#The GARCH part:

ylim<-c(min(na.omit(garch.e.t.VaR.10.save),na.omit(garch.e.t.ES.10.save),-
VaR.POT.vec.10.McNeil,-ES.POT.vec.10.McNeil,true.10),

max(na.omit(garch.e.t.VaR.10.save),na.omit(garch.e.t.ES.10.save),-VaR.POT.vec.10.McNeil, -
ES.POT.vec.10.McNeil,true.10))

plot(true.10,lty=1, main="Forecasts from EGARCH Student t model\ncombined with POT (ML
estimate of EVI)",ty="l",col="light blue",ylim=ylim,xaxt="n",xlab="",ylab="")

garch.e.t.VaR.10.add<-garch.e.t.VaR.10.save

garch.e.t.ES.10.add<-garch.e.t.ES.10.save

garch.e.t.VaR.10.add[indices.empty]<-" "

```

```

lines(garch.e.t.VaR.10.add,lty=2,col="red")

garch.e.t.ES.10.add[indices.empty]<-" "

lines(garch.e.t.ES.10.add,lty=4,col="green")

#legend(x="bottomright",legend=c("True returns","99% 10 day VaR(GARCH)","99% 10 day ES
(GARCH)","99% 10 day VaR (POT)","99% 10 day ES (POT)"),

#lty=c(1,2,4,2,4),lwd=c(1,1,1,2,2),col=c("light blue","red","green","darkred","darkgreen"))

#The EVT part:

VaR.POT.vec.10.McNeil.add<--VaR.POT.vec.10.McNeil

ES.POT.vec.10.McNeil.add<--ES.POT.vec.10.McNeil

VaR.POT.vec.10.McNeil.add[indices.work]<-" "

ES.POT.vec.10.McNeil.add[indices.work]<-" "

lines(VaR.POT.vec.10.McNeil.add,lty=2,lwd=2,col="darkred")

lines(ES.POT.vec.10.McNeil.add,lty=4,lwd=2,col="darkgreen")

axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.8)

#EGARCH t WITH Mkn

#The GARCH part:

ylim<-c(min(na.omit(garch.e.t.VaR.10.save),na.omit(garch.e.t.ES.10.save),-
VaR.POT.Mkn.vec.10.McNeil,-ES.POT.Mkn.vec.10.McNeil,true.10),

max(na.omit(garch.e.t.VaR.10.save),na.omit(garch.e.t.ES.10.save),-VaR.POT.Mkn.vec.10.McNeil, -
ES.POT.Mkn.vec.10.McNeil,true.10))

plot(true.10,lty=1, main="Forecasts from EGARCH Student t model\ncombined with POT (Moment
estimate of EVI)",ty="l",col="light blue",ylim=ylim,xaxt="n",xlab="",ylab="")

garch.e.t.VaR.10.add<-garch.e.t.VaR.10.save

garch.e.t.ES.10.add<-garch.e.t.ES.10.save

garch.e.t.VaR.10.add[indices.empty]<-" "

lines(garch.e.t.VaR.10.add,lty=2,col="red")

garch.e.t.ES.10.add[indices.empty]<-" "

lines(garch.e.t.ES.10.add,lty=4,col="green")

#legend(x="bottomright",legend=c("True returns","99% 10 day VaR(GARCH)","99% 10 day ES
(GARCH)","99% 10 day VaR (POT)","99% 10 day ES (POT)"),

#lty=c(1,2,4,2,4),lwd=c(1,1,1,2,2),col=c("light blue","red","green","darkred","darkgreen"))

#The EVT part:

VaR.POT.Mkn.vec.10.McNeil.add<--VaR.POT.Mkn.vec.10.McNeil

ES.POT.Mkn.vec.10.McNeil.add<--ES.POT.Mkn.vec.10.McNeil

VaR.POT.Mkn.vec.10.McNeil.add[indices.work]<-" "

ES.POT.Mkn.vec.10.McNeil.add[indices.work]<-" "

lines(VaR.POT.Mkn.vec.10.McNeil.add,lty=2,lwd=2,col="darkred")

```



```

lines(ES.POT.Mkn.vec.10.McNeil.add,lty=4,lwd=2,col="darkgreen")

axis(side=1,at=ax.vals.3,labels=dates.log.10[ax.vals.3,],cex.axis=0.8)

#Tables 8.8 and 8.9:

#ML+Normal:

#ML: VaR

true<-40.45

symm.VaR.ML.N<-
sum(na.omit(true.10[indices.work]<garch.n.VaR.10.save[indices.work]))+sum(na.omit(true.10[indices.empty]<(-VaR.POT.vec.10.McNeil)[indices.empty]))

gjr.VaR.ML.N<-
sum(na.omit(true.10[indices.work]<garch.gjr.VaR.10.save[indices.work]))+sum(na.omit(true.10[indices.empty]<(-VaR.POT.vec.10.McNeil)[indices.empty]))

e.VaR.ML.N<-
sum(na.omit(true.10[indices.work]<garch.e.n.VaR.10.save[indices.work]))+sum(na.omit(true.10[indices.empty]<(-VaR.POT.vec.10.McNeil)[indices.empty]))

#ML: ES

symm.ES.ML.N<-
sum(na.omit(true.10[indices.work]<garch.n.ES.10.save[indices.work]))+sum(na.omit(true.10[indices.empty]<(-ES.POT.vec.10.McNeil)[indices.empty]))

gjr.ES.ML.N<-
sum(na.omit(true.10[indices.work]<garch.gjr.ES.10.save[indices.work]))+sum(na.omit(true.10[indices.empty]<(-ES.POT.vec.10.McNeil)[indices.empty]))

e.ES.ML.N<-
sum(na.omit(true.10[indices.work]<garch.e.n.ES.10.save[indices.work]))+sum(na.omit(true.10[indices.empty]<(-ES.POT.vec.10.McNeil)[indices.empty]))

VaR.ML.N<-c(true,symm.VaR.ML.N,gjr.VaR.ML.N,e.VaR.ML.N)

ES.ML.N<-c(true,symm.ES.ML.N,gjr.ES.ML.N,e.ES.ML.N)

table.ML.N<-cbind(VaR.ML.N,ES.ML.N)

rownames(table.ML.N)<-c("Expected","Symmmetric","GJR GARCH","EGARCH")

colnames(table.ML.N)<-c("VaR","ES")

cat("Number of violations: GARCH N models combined with POT (ML Estimate of EVI)\n")

table.ML.N

#Moment+Normal:

#Moment+Normal:

#Moment: VaR

true<-40.45

```

```

symm.VaR.Mkn.N<-
sum(na.omit(true.10[indices.work]<garch.n.VaR.10.save[indices.work]))+sum(na.omit(true.10[indices.empty]<(-VaR.POT.Mkn.vec.10.McNeil)[indices.empty]))

gjr.VaR.Mkn.N<-
sum(na.omit(true.10[indices.work]<garch.gjr.VaR.10.save[indices.work]))+sum(na.omit(true.10[indices.empty]<(-VaR.POT.Mkn.vec.10.McNeil)[indices.empty]))

e.VaR.Mkn.N<-
sum(na.omit(true.10[indices.work]<garch.e.n.VaR.10.save[indices.work]))+sum(na.omit(true.10[indices.empty]<(-VaR.POT.Mkn.vec.10.McNeil)[indices.empty]))

#Moment: ES

symm.ES.Mkn.N<-
sum(na.omit(true.10[indices.work]<garch.n.ES.10.save[indices.work]))+sum(na.omit(true.10[indices.empty]<(-ES.POT.Mkn.vec.10.McNeil)[indices.empty]))

gjr.ES.Mkn.N<-
sum(na.omit(true.10[indices.work]<garch.gjr.ES.10.save[indices.work]))+sum(na.omit(true.10[indices.empty]<(-ES.POT.Mkn.vec.10.McNeil)[indices.empty]))

e.ES.Mkn.N<-
sum(na.omit(true.10[indices.work]<garch.e.n.ES.10.save[indices.work]))+sum(na.omit(true.10[indices.empty]<(-ES.POT.Mkn.vec.10.McNeil)[indices.empty]))

VaR.Mkn.N<-c(true,symm.VaR.Mkn.N,gjr.VaR.Mkn.N,e.VaR.Mkn.N)

ES.Mkn.N<-c(true,symm.ES.Mkn.N,gjr.ES.Mkn.N,e.ES.Mkn.N)

table.Mkn.N<-cbind(VaR.Mkn.N,ES.Mkn.N)

rownames(table.Mkn.N)<-c("Expected","Symmmetric","GJR GARCH","EGARCH")

colnames(table.Mkn.N)<-c("VaR","ES")

cat("Number of violations: GARCH N models combined with POT (Moment Estimate of EVI)\n")

table.Mkn.N

#ML+t:

#ML: VaR

true<-40.45

symm.VaR.ML.t<-
sum(na.omit(true.10[indices.work]<garch.t.VaR.10.save[indices.work]))+sum(na.omit(true.10[indices.empty]<(-VaR.POT.vec.10.McNeil)[indices.empty]))

gjr.VaR.ML.t<-
sum(na.omit(true.10[indices.work]<garch.gjr.t.VaR.10.save[indices.work]))+sum(na.omit(true.10[indices.empty]<(-VaR.POT.vec.10.McNeil)[indices.empty]))

e.VaR.ML.t<-
sum(na.omit(true.10[indices.work]<garch.e.t.VaR.10.save[indices.work]))+sum(na.omit(true.10[indices.empty]<(-VaR.POT.vec.10.McNeil)[indices.empty]))

#ML: ES

symm.ES.ML.t<-
sum(na.omit(true.10[indices.work]<garch.t.ES.10.save[indices.work]))+sum(na.omit(true.10[indices.empty]<(-ES.POT.vec.10.McNeil)[indices.empty]))

```

```

gjr.ES.ML.t<-
sum(na.omit(true.10[indices.work]<garch.gjr.t.ES.10.save[indices.work]))+sum(na.omit(true.10[indices.empty]<(-ES.POT.vec.10.McNeil)[indices.empty]))

e.ES.ML.t<-
sum(na.omit(true.10[indices.work]<garch.e.t.ES.10.save[indices.work]))+sum(na.omit(true.10[indices.empty]<(-ES.POT.vec.10.McNeil)[indices.empty]))

VaR.ML.t<-c(true,symm.VaR.ML.t,gjr.VaR.ML.t,e.VaR.ML.t)

ES.ML.t<-c(true,symm.ES.ML.t,gjr.ES.ML.t,e.ES.ML.t)

table.ML.t<-cbind(VaR.ML.t,ES.ML.t)

rownames(table.ML.t)<-c("Expected","Symmmetric","GJR GARCH","EGARCH")

colnames(table.ML.t)<-c("VaR","ES")

cat("Number of violations: GARCH Student t models combined with POT (ML Estimate of EVI)\n")

table.ML.t

#Moment+t:

#Moment: VaR

true<-40.45

symm.VaR.Mkn.t<-
sum(na.omit(true.10[indices.work]<garch.t.VaR.10.save[indices.work]))+sum(na.omit(true.10[indices.empty]<(-VaR.POT.Mkn.vec.10.McNeil)[indices.empty]))

gjr.VaR.Mkn.t<-
sum(na.omit(true.10[indices.work]<garch.gjr.t.VaR.10.save[indices.work]))+sum(na.omit(true.10[indices.empty]<(-VaR.POT.Mkn.vec.10.McNeil)[indices.empty]))

e.VaR.Mkn.t<-
sum(na.omit(true.10[indices.work]<garch.e.t.VaR.10.save[indices.work]))+sum(na.omit(true.10[indices.empty]<(-VaR.POT.Mkn.vec.10.McNeil)[indices.empty]))

#Moment: ES

symm.ES.Mkn.t<-
sum(na.omit(true.10[indices.work]<garch.t.ES.10.save[indices.work]))+sum(na.omit(true.10[indices.empty]<(-ES.POT.Mkn.vec.10.McNeil)[indices.empty]))

gjr.ES.Mkn.t<-
sum(na.omit(true.10[indices.work]<garch.gjr.t.ES.10.save[indices.work]))+sum(na.omit(true.10[indices.empty]<(-ES.POT.Mkn.vec.10.McNeil)[indices.empty]))

e.ES.Mkn.t<-
sum(na.omit(true.10[indices.work]<garch.e.t.ES.10.save[indices.work]))+sum(na.omit(true.10[indices.empty]<(-ES.POT.Mkn.vec.10.McNeil)[indices.empty]))

VaR.Mkn.t<-c(true,symm.VaR.Mkn.t,gjr.VaR.Mkn.t,e.VaR.Mkn.t)

ES.Mkn.t<-c(true,symm.ES.Mkn.t,gjr.ES.Mkn.t,e.ES.Mkn.t)

table.Mkn.t<-cbind(VaR.Mkn.t,ES.Mkn.t)

rownames(table.Mkn.t)<-c("Expected","Symmmetric","GJR GARCH","EGARCH")

```

```

colnames(table.Mkn.t)<-c("VaR","ES")

cat("Number of violations: GARCH Student t models combined with POT (Moment Estimate of
EVI)\n")

table.Mkn.t

#=====

#TWO STEP PROCEDURE (MCNEIL)

#Only do for VaR

#=====

#=====

#Scenario 1: Filter with i.i.d. n - then apply EVT - redo previous process
#Scenario 2: Filter with i.i.d. t - then apply EVT - redo previous process
#Scenario 3: Filter with Symm GARCH t - then apply EVT as far as possible
#=====

#=====

#Scenarios 1 and 2: Filter with i.i.d. n/t - then apply EVT - redo previous process
#=====

#(Change data.2 argument for i.i.d. n/t)

#Calculate the residuals=(true returns)-(forecasted VaR/ES): NB must be for one day first!!!

true.4.resids<-J580[251:4304,]

resids.iid.n.VaR<-true.4.resids-iid.n.VaR.save/sqrt(10) #need to convert back to 1 day
estimates

resids.iid.t.VaR<-true.4.resids-iid.t.VaR.save/sqrt(10) #need to convert back to 1 day
estimates

#resids.iid.n.VaR.10 ; resids.iid.t.VaR.10: from excel workbook: "Resids vir two step" (dink
nie ons gebruik dit nie)

#true.10.2: from excel workbook: "Resids vir two step"

#Notes:

#Remember the 10 day estimates were originally for day 260-4313, i.e. 4054 days.

#But converting back to 1 day estimates, so it must be for 4054 days starting from day 251

#So true.4.resids is the J580[251:4304,] yielding 4054 returns

#So not using the last 9 return values.

```

```

#POT approach with ML estimates

#A threshold of 20% or 50 observations will be used.

data.2<-as.data.frame(resids.iid.n.VaR) #CHANGE THIS TO resids.iid.t.VaR and re-run everything
for t

#First estimate parameters of GPD using ML and moving data window of 250 days.

#(Next steps takes a few minutes to run)

library(ismev) #adjusted gpd.fit to gpd.fit.own- give only MLEs in vector
gamma.vec.2<-rep(0,3804)
sigma.vec.2<-rep(0,3804)
M.k.n.vec.2<-rep(0,3804)

k<-50

for (i in 1:3804){
data.x.2<-data.2[i:(i+250-1),]
#for MLEs:
gamma.vec.2[i]<-gpd.fit.own(data.x.2,threshold=thresholds[4],npy=250)[2]
sigma.vec.2[i]<-gpd.fit.own(data.x.2,threshold=thresholds[4],npy=250)[1]
#for Mkn:
data.ordered.x.2<-sort(data.x.2,decreasing=FALSE)
gamma.hill.2<-mean(log(data.ordered.x.2[(250-k+1):250]))-log(data.ordered.x.2[(250-k)])
gamma.hill.2.2<- mean((log(data.ordered.x.2[(250-k+1):250]))-log(data.ordered.x.2[(250-k)]))^2)
M.k.n.vec.2[i]<-gamma.hill.2 + 1 - 0.5*(1-(gamma.hill.2)^2/gamma.hill.2.2)^(-1)
}

gamma.vec.save.2<-gamma.vec.2
sigma.vec.save.2<-sigma.vec.2
M.k.n.vec.save.2<-M.k.n.vec.2

plot(M.k.n.vec.2,ty="l",ylim=c(min(M.k.n.vec.2,gamma.vec.2),max(M.k.n.vec.2,gamma.vec.2)),lty=
1,col="blue",

main="Estimators of the EVI for moving data window of length 250 days",ylab="",xlab="")

lines(gamma.vec.2,lty=2,col="red")

legend(legend=c("Moment Estimator","ML estimate of GPD
gamma"),x="topright",col=c("blue","red"),lty=1:2)

```

```

plot(sigma.vec.2,xlab="",ylab="",main="ML estimate of GPD sigma",ty="l")

#The POT method for constructing VaR and ES (using ML estimates)
#This is for 99%
k<-50
Nu<-k
alpha<-0.01
VaR.POT.vec.2<-rep(0,3804) #The 1 day forecasts (250 data window)
ES.POT.vec.2<-rep(0,3804)
u.vec.2<-rep(0,3804) #Thresholds for each of the 250 day windows

for (i in 1:3804){
  data.x.2<-data.2[i:(i+250-1),]
  data.ordered.x.2<-sort(data.x.2,decreasing=FALSE)
  u.2<-data.ordered.x.2[250-k] #The threshold
  u.vec.2[i]<-u.2
  sigma.2<-sigma.vec.2[i]
  gamma.2<-gamma.vec.2[i]
  VaR.POT.2<-u.2+(sigma.2/gamma.2)*(((250*alpha)/Nu)^(-gamma.2)-1)
  VaR.POT.vec.2[i]<-u.2+(sigma.2/gamma.2)*(((250*alpha)/Nu)^(-gamma.2)-1)
  ES.POT.vec.2[i]<-VaR.POT.2/(1-gamma.2)+(sigma.2-gamma.2*u.2)/(1-gamma.2)
}

VaR.POT.vec.save.2<-VaR.POT.vec.2
ES.POT.vec.save.2<-ES.POT.vec.2

#Now using Mkn
k<-50
Nu<-k
alpha<-0.01
VaR.POT.Mkn.vec.2<-rep(0,3804) #The 1 day forecasts (250 data window)
ES.POT.Mkn.vec.2<-rep(0,3804)

for (i in 1:3804){
  data.x.2<-data.2[i:(i+250-1),]
  data.ordered.x.2<-sort(data.x.2,decreasing=FALSE)
  u.2<-data.ordered.x.2[250-k] #The threshold

```

```

sigma.2<-sigma.vec.2[i]
gamma.2<-M.k.n.vec.2[i]
VaR.POT.Mkn.vec.2[i]<-u.2+(sigma.2/gamma.2)*(((250*alpha)/Nu)^(-gamma.2)-1)
VaR.POT.2<-u.2+(sigma.2/gamma.2)*(((250*alpha)/Nu)^(-gamma.2)-1)
ES.POT.Mkn.vec.2[i]<-VaR.POT.2/(1-gamma.2)+(sigma.2-gamma.2*u.2)/(1-gamma.2)
}

VaR.POT.Mkn.vec.save.2<-VaR.POT.Mkn.vec.2
ES.POT.Mkn.vec.save.2<-ES.POT.Mkn.vec.2
u.vec.save.2<-u.vec.2

#The plots: (1 DAY ESTIMATES)
#POT and ML:
true.2<- true.4.resids[251:length(true.4.resids)]

#The resids have length 4054; there are 3804=4054-250+1 return values that can be estimated
(from day 251 onwards)

plot(true.2,ty="l",col="lightblue",main="1 day POT VaR and ES estimates (ML estimate of
EVI)",xlab="",ylab="",
ylim=c(min(true.2,-VaR.POT.vec.2,-ES.POT.vec.2),max(true.2,-VaR.POT.vec.2,-ES.POT.vec.2)))
lines(-VaR.POT.vec.2,col="red",lty=2)
lines(-ES.POT.vec.2,col="green",lty=3)

legend(legend=c("True daily log returns","99% 1 day VaR","99% 1 day
ES"),x="bottomright",col=c("lightblue","red","green"),lty=1:3)

#POT and Mkn:
plot(true.2,ty="l",col="lightblue",main="1 day POT VaR and ES estimates (Moment estimate of
EVI)",xlab="",ylab="",
ylim=c(min(true.2,-VaR.POT.Mkn.vec.2,-ES.POT.Mkn.vec.2),max(true.2,-VaR.POT.Mkn.vec.2,-
ES.POT.Mkn.vec.2)))
lines(-VaR.POT.Mkn.vec.2,col="red",lty=2)
lines(-ES.POT.Mkn.vec.2,col="green",lty=3)

legend(legend=c("True daily log returns","99% 1 day VaR","99% 1 day
ES"),x="bottomright",col=c("lightblue","red","green"),lty=1:3)

#ONLY VAR:
#POT and ML:
true.2<- true.4.resids[251:length(true.4.resids)]

#The resids have length 4054; there are 3804=4054-250+1 return values that can be estimated
(from day 251 onwards)

plot(true.2,ty="l",col="lightblue",main="1 day POT VaR and ES estimates (ML estimate of
EVI)",xlab="",ylab="",

```

```

ylim=c(min(true.2,-VaR.POT.vec.2),max(true.2,-VaR.POT.vec.2))

lines(-VaR.POT.vec.2,col="red",lty=2)

legend(legend=c("True daily log returns","99% 1 day
VaR"),x="bottomright",col=c("lightblue","red"),lty=1:2)

#POT and Mkn:

plot(true.2,ty="l",col="lightblue",main="1 day POT VaR and ES estimates (Moment estimate of
EVI)",xlab="",ylab="",

ylim=c(min(true.2,-VaR.POT.Mkn.vec.2),max(true.2,-VaR.POT.Mkn.vec.2))

lines(-VaR.POT.Mkn.vec.2,col="red",lty=2)

legend(legend=c("True daily log returns","99% 1 day
VaR"),x="bottomright",col=c("lightblue","red"),lty=1:3)

#Table of exceedances:

exceed.VaR.POT.1.2<-sum(true.2<(-VaR.POT.vec.2))

exceed.ES.POT.1.2<-sum(true.2<(-ES.POT.vec.2))

exceed.VaR.POT.Mkn.1.2<-sum(true.2<(-VaR.POT.Mkn.vec.2))

exceed.ES.POT.Mkn.1.2<-sum(true.2<(-ES.POT.Mkn.vec.2))

table.x.POT.1.2<-
rbind(c(38.04,38.04),c(exceed.VaR.POT.1.2,exceed.ES.POT.1.2),c(exceed.VaR.POT.Mkn.1.2,exceed.E
S.POT.Mkn.1.2))

colnames(table.x.POT.1.2)<-c("VaR","ES")

rownames(table.x.POT.1.2)<-c("Expected","POT ML estimator","POT Moment estimator")

table.x.POT.1.2

#####10 DAY ESTIMATES: SCALING#####

#"Save" values later on

#To convert the one day VaR to a 10 day VaR, use the alpha root rule:

#VaR(10)=VaR(1)(10)^(gamma)

#E.g. 10 day ahead forecast made on day 251 for day 260 is...

y.2<-3795 #since true.10.2 contains 3795 returns

VaR.POT.vec.10.2<-VaR.POT.vec.2[1:y.2]*(10)^(gamma.vec.2[1:y.2])

VaR.POT.Mkn.vec.10.2<-VaR.POT.Mkn.vec.2[1:y.2]*(10)^(M.k.n.vec.2[1:y.2])

ES.POT.vec.10.2<-VaR.POT.vec.10.2/(1-gamma.vec.save.2[1:y.2])+(sigma.vec.save.2[1:y.2]-
gamma.vec.save.2[1:y.2]*u.vec.save.2[1:y.2])/(1-gamma.vec.save.2[1:y.2])

```



```

ES.POT.Mkn.vec.10.2<-VaR.POT.Mkn.vec.10.2/(1-
M.k.n.vec.save.2[1:y.2])+(sigma.vec.save.2[1:y.2]-
M.k.n.vec.save.2[1:y.2]*u.vec.save.2[1:y.2])/(1-M.k.n.vec.save.2[1:y.2])

#Conversion using square root of time rule:

y.2<-3795      #since true.10.2 contains 3795 returns

VaR.POT.vec.10.2<-VaR.POT.vec.2[1:y.2]*(10)^(0.5)

VaR.POT.Mkn.vec.10.2<-VaR.POT.Mkn.vec.2[1:y.2]*(10)^(0.5)

ES.POT.vec.10.2<-VaR.POT.vec.10.2/(1-gamma.vec.save.2[1:y.2])+(sigma.vec.save.2[1:y.2]-
gamma.vec.save.2[1:y.2]*u.vec.save.2[1:y.2])/(1-gamma.vec.save.2[1:y.2])

ES.POT.Mkn.vec.10.2<-VaR.POT.Mkn.vec.10.2/(1-
M.k.n.vec.save.2[1:y.2])+(sigma.vec.save.2[1:y.2]-
M.k.n.vec.save.2[1:y.2]*u.vec.save.2[1:y.2])/(1-M.k.n.vec.save.2[1:y.2])

#The empirically established scaling laws:

#First establish the historical distribution and the percentiles.

returns.2<-as.data.frame(resids.iid.n.VaR) #CHANGE IF WORKING WITH t DISTRIBUTION

hist.sd.2<-rep(0,3804) #The hist sd's for ALL the 250 day window periods

for (i in 1:length(hist.sd.2)){
hist.sd.2[i]<-sd(returns.2[i:(250+i-1)],)
}

hist.quantiles.5.2<-quantile(hist.sd.2,probs=0.05)

hist.quantiles.50.2<-quantile(hist.sd.2,probs=0.5)

hist.quantiles.95.2<-quantile(hist.sd.2,probs=0.95)

hist.quantiles.2<-c(hist.quantiles.5.2,hist.quantiles.50.2,hist.quantiles.95.2)

a.2<-hist.quantiles.2[[1]]

b.2<-hist.quantiles.2[[2]]

c.2<-hist.quantiles.2[[3]]

slope1.2<-(0.65-0.59)/(a.2-b.2)

intercept1.2<-0.65-((0.65-0.59)/(a.2-b.2))*a.2

slope2.2<-(0.59-0.47)/(b.2-c.2)

intercept2.2<-0.59-((0.59-0.47)/(b.2-c.2))*b.2

#plot the function used for the interpolation:

plot(x=c(min(hist.sd.2),a.2,b.2,c.2,max(hist.sd.2)),y=c(0.65,0.65,0.59,0.47,0.47),ty="b",
main="Linear interpolation of scaling factors",xlab="Standard deviation",ylab="Scaling
factor")

text(x=c(a.2,b.2,c.2),y=c(0.65,0.59,0.47),labels=c("a","b","c"),pos=4)

```

```

#NB: Fill in values of a.2,b.2,c.2!!!
legend(x="topright",legend=c("a=(0.02405,0.65)","b=(0.03749,0.59)","c=(0.10464,0.47)"))

VaR.POT.vec.10.2<-rep(0,3795)
ES.POT.vec.10.2<-rep(0,3795)
VaR.POT.Mkn.vec.10.2<-rep(0,3795)
ES.POT.Mkn.vec.10.2<-rep(0,3795)
scale.vec.2<-rep(0,3795)
for (i in 1:length(VaR.POT.vec.10.2)){
  if(hist.sd.2[i]<a.2) {sc.2<-0.65}
  if (hist.sd.2[i]==a.2) {sc.2<-0.65}
  if((a.2<hist.sd.2[i])&(hist.sd.2[i]<b.2)) {sc.2<-slope1.2*hist.sd.2[i]+intercept1.2}
  if(hist.sd.2[i]==b.2) {sc.2<-0.59}
  if((b.2<hist.sd.2[i])&(hist.sd.2[i]<c.2)) {sc.2<-slope2.2*hist.sd.2[i]+intercept2.2}
  if (hist.sd.2[i]==c.2) {sc.2<-0.47}
  if(hist.sd.2[i]>=c.2) {sc.2<-0.47}
  scale.2<-sc.2
  scale.vec.2[i]<-sc.2
  VaR.POT.vec.10.i.2<-VaR.POT.vec.save.2[i]*(10)^scale.2
  VaR.POT.vec.10.2[i]<-VaR.POT.vec.10.i.2
  VaR.POT.Mkn.vec.10.i.2<-VaR.POT.Mkn.vec.save.2[i]*(10)^scale.2
  VaR.POT.Mkn.vec.10.2[i]<-VaR.POT.Mkn.vec.10.i.2
  ES.POT.vec.10.2[i]<-VaR.POT.vec.10.i.2/(1-gamma.vec.save.2[i])+(sigma.vec.save.2[i]-
gamma.vec.save.2[i]*u.vec.save.2[i])/(1-gamma.vec.save.2[i])
  ES.POT.Mkn.vec.10.2[i]<-VaR.POT.Mkn.vec.10.i.2/(1-M.k.n.vec.save.2[i])+(sigma.vec.save.2[i]-
M.k.n.vec.save.2[i]*u.vec.save.2[i])/(1-M.k.n.vec.save.2[i])
}
plot(scale.vec.2,cex=0.5,main="Scaling factors used",ylab="Scaling factor")

#Saving the values of ES and VaR for each of the 3 methods (and the ML and Mkn estimates):
#Square root of time rule:
VaR.POT.vec.10.SquareRoot.2<-VaR.POT.vec.10.2
ES.POT.vec.10.SquareRoot.2<-ES.POT.vec.10.2
VaR.POT.Mkn.vec.10.SquareRoot.2<-VaR.POT.Mkn.vec.10.2
ES.POT.Mkn.vec.10.SquareRoot.2<-ES.POT.Mkn.vec.10.2

#Alpha root rule:

```

```

VaR.POT.vec.10.AlphaRoot.2<-VaR.POT.vec.10.2
ES.POT.vec.10.AlphaRoot.2<-ES.POT.vec.10.2
VaR.POT.Mkn.vec.10.AlphaRoot.2<-VaR.POT.Mkn.vec.10.2
ES.POT.Mkn.vec.10.AlphaRoot.2<-ES.POT.Mkn.vec.10.2

#Empirically established scaling laws (McNeil):
VaR.POT.vec.10.McNeil.2<-VaR.POT.vec.10.2
ES.POT.vec.10.McNeil.2<-ES.POT.vec.10.2
VaR.POT.Mkn.vec.10.McNeil.2<-VaR.POT.Mkn.vec.10.2
ES.POT.Mkn.vec.10.McNeil.2<-ES.POT.Mkn.vec.10.2

#dates.5<-read.table(file="clipboard")
ax.vals.5<-c(1,floor(3795*spacings))

#THE PLOTS: (Re-run for each method!)
#For the ML estimate of the EVI:
plot(true.10.2,ty="l",col="lightblue",main="10 day POT VaR and ES estimates (ML estimate of
EVI)",xlab="",ylab="",
ylim=c(min(true.10.2,-VaR.POT.vec.10.2,-ES.POT.vec.10.2),max(true.10.2,-VaR.POT.vec.10.2,-
ES.POT.vec.10.2)),xaxt="n")
lines(-VaR.POT.vec.10.2,col="red",lty=2)
lines(-ES.POT.vec.10.2,col="green",lty=3)
legend(legend=c("True 10 day returns","99% 10 day VaR","99% 10 day
ES"),x="topright",col=c("lightblue","red","green"),lty=1:3)
axis(side=1,at=ax.vals.5,labels=dates.5[ax.vals.5,],cex.axis=0.8)

#For the Mkn estimate of the EVI:
plot(true.10.2,ty="l",col="lightblue",main="10 day POT VaR and ES estimates (Moment of
EVI)",xlab="",ylab="",
ylim=c(min(true.10.2,-VaR.POT.vec.10.2,-VaR.POT.Mkn.vec.10.2,-
ES.POT.Mkn.vec.10.2),max(true.10.2,-VaR.POT.Mkn.vec.10.2,-ES.POT.Mkn.vec.10.2)))
lines(-VaR.POT.Mkn.vec.10.2,col="red",lty=2)
lines(-ES.POT.Mkn.vec.10.2,col="green",lty=3)
legend(legend=c("True 10 day returns","99% 10 day VaR","99% 10 day
ES"),x="topright",col=c("lightblue","red","green"),lty=1:3)

#ONLY VAR:
#For the ML estimate of the EVI:

```

```

plot(true.10.2,ty="l",col="lightblue",main="10 day POT VaR and ES estimates (ML estimate of
EVI)",xlab="",ylab="",

ylim=c(min(true.10.2,-VaR.POT.vec.10.2),max(true.10.2,-VaR.POT.vec.10.2)),xaxt="n")

lines(-VaR.POT.vec.10.2,col="red",lty=2)

legend(legend=c("True 10 day returns","99% 10 day
VaR"),x="topright",col=c("lightblue","red"),lty=1:2)

axis(side=1,at=ax.vals.5,labels=dates.5[ax.vals.5,],cex.axis=0.8)

#For the Mkn estimate of the EVI:

plot(true.10.2,ty="l",col="lightblue",main="10 day POT VaR and ES estimates (Moment of
EVI)",xlab="",ylab="",

ylim=c(min(true.10.2,-VaR.POT.vec.10.2,-VaR.POT.Mkn.vec.10.2),max(true.10.2,-
VaR.POT.Mkn.vec.10.2)),xaxt="n")

lines(-VaR.POT.Mkn.vec.10.2,col="red",lty=2)

legend(legend=c("True 10 day returns","99% 10 day
VaR"),x="topright",col=c("lightblue","red"),lty=1:2)

axis(side=1,at=ax.vals.5,labels=dates.5[ax.vals.5,],cex.axis=0.8)

#The tables (also re-run for each method)

violations.VaR.POT.2<-sum(true.10.2<(-VaR.POT.vec.10.2))

violations.ES.POT.2<-sum(true.10.2<(-ES.POT.vec.10.2))

table.2<-rbind(37.95,violations.VaR.POT.2,violations.ES.POT.2)

colnames(table.2)<-"Number of violations"

rownames(table.2)<-c("Expected","VaR","ES")

table.2

violations.VaR.POT.Mkn.2<-sum(true.10.2<(-VaR.POT.Mkn.vec.10.2[1:length(true.10.2)]))

violations.ES.POT.Mkn.2<-sum(true.10.2<(-ES.POT.Mkn.vec.10.2[1:length(true.10.2)]))

table.2<-rbind(37.95,violations.VaR.POT.Mkn.2,violations.ES.POT.Mkn.2)

colnames(table.2)<-"Number of violations"

rownames(table.2)<-c("Expected","VaR","ES")

table.2

#Comparative plots on same axis:

#VaR+ML for 3 different methods:

```

```

plot(true.10.2,ty="l",col="lightblue",main="99% 10 day POT VaR estimates (ML estimate of
EVI)",xlab="",ylab="",

ylim=c(min(true.10.2,-VaR.POT.vec.10.SquareRoot.2,-VaR.POT.vec.10.AlphaRoot.2,-
VaR.POT.vec.10.McNeil.2),

max(true.10.2,-VaR.POT.vec.10.SquareRoot.2,-VaR.POT.vec.10.AlphaRoot.2,-
VaR.POT.vec.10.McNeil.2)))

lines(-VaR.POT.vec.10.SquareRoot.2,col="red",lty=2)

lines(-VaR.POT.vec.10.AlphaRoot.2,col="darkgreen",lty=3)

lines(-VaR.POT.vec.10.McNeil.2,col="purple",lty=4)

legend(legend=c("True 10 day log returns","Square Root","Alpha Root","Empirical scaling
laws"),x="bottomright",col=c("lightblue","red","darkgreen","purple"),lty=1:4)

#VaR+Mkn for 3 different methods:

plot(true.10.2,ty="l",col="lightblue",main="99% 10 day POT VaR estimates (Moment estimate of
EVI)",xlab="",ylab="",

ylim=c(min(true.10.2,-VaR.POT.Mkn.vec.10.SquareRoot.2,-VaR.POT.Mkn.vec.10.AlphaRoot.2,-
VaR.POT.Mkn.vec.10.McNeil.2),

max(true.10.2,-VaR.POT.Mkn.vec.10.SquareRoot.2,-VaR.POT.Mkn.vec.10.AlphaRoot.2,-
VaR.POT.Mkn.vec.10.McNeil.2)))

lines(-VaR.POT.Mkn.vec.10.SquareRoot.2,col="red",lty=2)

lines(-VaR.POT.Mkn.vec.10.AlphaRoot.2,col="darkgreen",lty=3)

lines(-VaR.POT.Mkn.vec.10.McNeil.2,col="purple",lty=4)

legend(legend=c("True 10 day log returns","Square Root","Alpha Root","Empirical scaling
laws"),x="bottomright",col=c("lightblue","red","darkgreen","purple"),lty=1:4)

#ES+ML for 3 different methods:

plot(true.10.2,ty="l",col="lightblue",main="99% 10 day POT ES estimates (ML estimate of
EVI)",xlab="",ylab="",

ylim=c(min(true.10.2,-ES.POT.vec.10.SquareRoot.2,-ES.POT.vec.10.AlphaRoot.2,-
ES.POT.vec.10.McNeil.2),

max(true.10.2,-ES.POT.vec.10.SquareRoot.2,-ES.POT.vec.10.AlphaRoot.2,-
ES.POT.vec.10.McNeil.2)))

lines(-ES.POT.vec.10.SquareRoot.2,col="red",lty=2)

lines(-ES.POT.vec.10.AlphaRoot.2,col="darkgreen",lty=3)

lines(-ES.POT.vec.10.McNeil.2,col="purple",lty=4)

legend(legend=c("True 10 day log returns","Square Root","Alpha Root","Empirical scaling
laws"),x="bottomright",col=c("lightblue","red","darkgreen","purple"),lty=1:4)

#ES+Mkn for 3 different methods:

plot(true.10.2,ty="l",col="lightblue",main="99% 10 day POT ES estimates (Moment estimate of
EVI)",xlab="",ylab="",

```

```

ylim=c(min(true.10.2,-ES.POT.Mkn.vec.10.SquareRoot.2,-ES.POT.Mkn.vec.10.AlphaRoot.2,-
ES.POT.Mkn.vec.10.McNeil.2),
max(true.10.2,-ES.POT.Mkn.vec.10.SquareRoot.2,-ES.POT.Mkn.vec.10.AlphaRoot.2,-
ES.POT.Mkn.vec.10.McNeil.2))

lines(-ES.POT.Mkn.vec.10.SquareRoot.2,col="red",lty=2)

lines(-ES.POT.Mkn.vec.10.AlphaRoot.2,col="darkgreen",lty=3)

lines(-ES.POT.Mkn.vec.10.McNeil.2,col="purple",lty=4)

legend(legend=c("True 10 day log returns","Square Root","Alpha Root","Empirical scaling
laws"),x="bottomright",col=c("lightblue","red","darkgreen","purple"),lty=1:4)

#####

#Scenario 3: Filter with Symm GARCH t - then apply EVT - redo previous process

#####

#The data will be fragmented:

#period 1: 1 to 444

#period 2: 695 to 740

#period 3: 1100 to 1460

#period 4: 1736 to 1842

#period 5: 3086 to 3629

#period 6: 3907 to 3917

#Each period will be handled separately. This is mainly for the application of the
#empirical scaling law. For the latter, only consider the distribution in each fragment;
#not using information from other fragments.

#IN THE CODE BELOW; CHANGE THE FOLLOWING DUMMY VARIABLES TO MOVE TO ANOTHER PERIOD:

#ll : the length of the period (number of data points): change to ll.1, ll.2, etc.

#resids.garch.t.VaR should be changed to resids.garch.t.VaR.1,resids.garch.t.VaR.2, etc.

#Dummy variables for the different periods:

#(Change data.2 argument for i.i.d. n/t)

#Calculate the residuals=(true returns)-(forecasted VaR/ES): NB must be for one day first!!!

true.4.resids<-J580[251:4304,]

resids.garch.t.VaR.pre<-true.4.resids-garch.t.VaR.save #One day estimates

#true.10.2: from excel workbook: "Resids vir two step"

#Notes:

```

```
#Remember the 10 day estimates were originally for day 260-4313, i.e. 4054 days.
#But converting back to 1 day estimates, so it must be for 4054 days starting from day 251
#So true.4.resids is the J580[251:4304,] yielding 4054 returns
#So not using the last 9 return values.

#Dummy variables for the different periods:
resids.garch.t.VaR.1<-resids.garch.t.VaR.pre[1:444]
resids.garch.t.VaR.2<-resids.garch.t.VaR.pre[695:740]
resids.garch.t.VaR.3<-resids.garch.t.VaR.pre[1100:1460]
resids.garch.t.VaR.4<-resids.garch.t.VaR.pre[1736:1842]
resids.garch.t.VaR.5<-resids.garch.t.VaR.pre[3086:3629]
resids.garch.t.VaR.6<-resids.garch.t.VaR.pre[3907:3917]
l1.1<-length(resids.garch.t.VaR.1)
l1.2<-length(resids.garch.t.VaR.2)
l1.3<-length(resids.garch.t.VaR.3)
l1.4<-length(resids.garch.t.VaR.4)
l1.5<-length(resids.garch.t.VaR.5)
l1.6<-length(resids.garch.t.VaR.6)

#IMPORTANT: ONLY PERIODS 1,3 AND 5 ARE LONGER THAN 250 DAYS
#THEREFORE ONLY THESE PERIODS CAN BE USED FOR ROLLING 250 DAY ESTIMATIONS

#POT approach with ML estimates
#A threshold of 20% or 50 observations will be used.

#CHANGE THE FOLLOWING DUMMY VARIABLE AS RELEVANT:
resids.garch.t.VaR<-resids.garch.t.VaR.5
data.2<-as.data.frame(resids.garch.t.VaR)
l1<-l1.5

#First estimate parameters of GPD using ML and moving data window of 250 days.

#(Next steps takes a few minutes to run)
library(ismev) #adjusted gpd.fit to gpd.fit.own- give only MLEs in vector
gamma.vec.2<-rep(0,(l1-250))
sigma.vec.2<-rep(0,(l1-250))
M.k.n.vec.2<-rep(0,(l1-250))
```

```

k<-50
for (i in 1:(11-250)){
data.x.2<-data.2[i:(i+250-1),]
#for MLEs:
gamma.vec.2[i]<-gpd.fit.own(data.x.2,threshold=thresholds[4],npy=250)[2]
sigma.vec.2[i]<-gpd.fit.own(data.x.2,threshold=thresholds[4],npy=250)[1]
#for Mkn:
data.ordered.x.2<-sort(data.x.2,decreasing=FALSE)
gamma.hill.2<-mean(log(data.ordered.x.2[(250-k+1):250]))-log(data.ordered.x.2[(250-k)])
gamma.hill.2.2<- mean((log(data.ordered.x.2[(250-k+1):250]))-log(data.ordered.x.2[(250-k)]))^2)
M.k.n.vec.2[i]<-gamma.hill.2 + 1 - 0.5*(1-(gamma.hill.2)^2/gamma.hill.2.2)^(-1)
}

gamma.vec.save.2<-gamma.vec.2
sigma.vec.save.2<-sigma.vec.2
M.k.n.vec.save.2<-M.k.n.vec.2

#The POT method for constructing VaR and ES (using ML estimates)
#This is for 99%
k<-50
Nu<-k
alpha<-0.01
VaR.POT.vec.2<-rep(0,(11-250)) #The 1 day forecasts (250 data window)
ES.POT.vec.2<-rep(0,(11-250))
u.vec.2<-rep(0,(11-250)) #Thresholds for each of the 250 day windows

for (i in 1:(11-250)){
  data.x.2<-data.2[i:(i+250-1),]
  data.ordered.x.2<-sort(data.x.2,decreasing=FALSE)
  u.2<-data.ordered.x.2[250-k] #The threshold
  u.vec.2[i]<-u.2
  sigma.2<-sigma.vec.2[i]
  gamma.2<-gamma.vec.2[i]
  VaR.POT.2<-u.2+(sigma.2/gamma.2)*(((250*alpha)/Nu)^(-gamma.2)-1)
  VaR.POT.vec.2[i]<-u.2+(sigma.2/gamma.2)*(((250*alpha)/Nu)^(-gamma.2)-1)
  ES.POT.vec.2[i]<-VaR.POT.2/(1-gamma.2)+(sigma.2-gamma.2*u.2)/(1-gamma.2)
}

```



```

    }

VaR.POT.vec.save.2<-VaR.POT.vec.2
ES.POT.vec.save.2<-ES.POT.vec.2

#Now using Mkn
k<-50
Nu<-k
alpha<-0.01
VaR.POT.Mkn.vec.2<-rep(0,(11-250)) #The 1 day forecasts (250 data window)
ES.POT.Mkn.vec.2<-rep(0,(11-250))

for (i in 1:(11-250)){
    data.x.2<-data.2[i:(i+250-1),]
    data.ordered.x.2<-sort(data.x.2,decreasing=FALSE)
    u.2<-data.ordered.x.2[250-k] #The threshold
    sigma.2<-sigma.vec.2[i]
    gamma.2<-M.k.n.vec.2[i]
    VaR.POT.Mkn.vec.2[i]<-u.2+(sigma.2/gamma.2)*(((250*alpha)/Nu)^(-gamma.2)-1)
    VaR.POT.2<-u.2+(sigma.2/gamma.2)*(((250*alpha)/Nu)^(-gamma.2)-1)
    ES.POT.Mkn.vec.2[i]<-VaR.POT.2/(1-gamma.2)+(sigma.2-gamma.2*u.2)/(1-gamma.2)
}

VaR.POT.Mkn.vec.save.2<-VaR.POT.Mkn.vec.2
ES.POT.Mkn.vec.save.2<-ES.POT.Mkn.vec.2
u.vec.save.2<-u.vec.2

#####10 DAY ESTIMATES: SCALING#####

#"Save" values later on

#To convert the one day VaR to a 10 day VaR, use the alpha root rule:
#VaR(10)=VaR(1)(10)^(gamma)
#E.g. 10 day ahead forecast made on day 251 for day 260 is....
y.2<-(11-260+1)
VaR.POT.vec.10.2<-VaR.POT.vec.2[1:y.2]*(10)^(gamma.vec.2[1:y.2])

```

```

VaR.POT.Mkn.vec.10.2<-VaR.POT.Mkn.vec.2[1:y.2]*(10)^(M.k.n.vec.2[1:y.2])

ES.POT.vec.10.2<-VaR.POT.vec.10.2/(1-gamma.vec.save.2[1:y.2])+(sigma.vec.save.2[1:y.2]-
gamma.vec.save.2[1:y.2]*u.vec.save.2[1:y.2])/(1-gamma.vec.save.2[1:y.2])

ES.POT.Mkn.vec.10.2<-VaR.POT.Mkn.vec.10.2/(1-
M.k.n.vec.save.2[1:y.2])+(sigma.vec.save.2[1:y.2]-
M.k.n.vec.save.2[1:y.2]*u.vec.save.2[1:y.2])/(1-M.k.n.vec.save.2[1:y.2])

#Conversion using square root of time rule: #DONE#

y.2<-(11-260+1)      #since true.10.2 contains 3795 returns

VaR.POT.vec.10.2<-VaR.POT.vec.2[1:y.2]*(10)^(0.5)

VaR.POT.Mkn.vec.10.2<-VaR.POT.Mkn.vec.2[1:y.2]*(10)^(0.5)

ES.POT.vec.10.2<-VaR.POT.vec.10.2/(1-gamma.vec.save.2[1:y.2])+(sigma.vec.save.2[1:y.2]-
gamma.vec.save.2[1:y.2]*u.vec.save.2[1:y.2])/(1-gamma.vec.save.2[1:y.2])

ES.POT.Mkn.vec.10.2<-VaR.POT.Mkn.vec.10.2/(1-
M.k.n.vec.save.2[1:y.2])+(sigma.vec.save.2[1:y.2]-
M.k.n.vec.save.2[1:y.2]*u.vec.save.2[1:y.2])/(1-M.k.n.vec.save.2[1:y.2])

#The empirically established scaling laws:

#First establish the historical distribution and the percentiles.

returns.2<-as.data.frame(resids.garch.t.VaR)

hist.sd.2<-rep(0,(11-250)) #The hist sd's for ALL the 250 day window periods
for (i in 1:length(hist.sd.2)){
hist.sd.2[i]<-sd(returns.2[i:(250+i-1),])
}

hist.quantiles.5.2<-quantile(hist.sd.2,probs=0.05)
hist.quantiles.50.2<-quantile(hist.sd.2,probs=0.5)
hist.quantiles.95.2<-quantile(hist.sd.2,probs=0.95)

hist.quantiles.2<-c(hist.quantiles.5.2,hist.quantiles.50.2,hist.quantiles.95.2)

a.2<-hist.quantiles.2[[1]]
b.2<-hist.quantiles.2[[2]]
c.2<-hist.quantiles.2[[3]]

slope1.2<-(0.65-0.59)/(a.2-b.2)

intercept1.2<-0.65-((0.65-0.59)/(a.2-b.2))*a.2

slope2.2<-(0.59-0.47)/(b.2-c.2)

intercept2.2<-0.59-((0.59-0.47)/(b.2-c.2))*b.2

VaR.POT.vec.10.2<-rep(0,(11-260+1))

ES.POT.vec.10.2<-rep(0,(11-260+1))

```

```

VaR.POT.Mkn.vec.10.2<-rep(0,(11-260+1))
ES.POT.Mkn.vec.10.2<-rep(0,(11-260+1))
scale.vec.2<-rep(0,(11-260+1))
for (i in 1:length(VaR.POT.vec.10.2)){
if(hist.sd.2[i]<a.2) {sc.2<-0.65}
if (hist.sd.2[i]==a.2) {sc.2<-0.65}
if((a.2<hist.sd.2[i])&(hist.sd.2[i]<b.2)) {sc.2<-slope1.2*hist.sd.2[i]+intercept1.2}
if(hist.sd.2[i]==b.2) {sc.2<-0.59}
if((b.2<hist.sd.2[i])&(hist.sd.2[i]<c.2)) {sc.2<-slope2.2*hist.sd.2[i]+intercept2.2}
if (hist.sd.2[i]==c.2) {sc.2<-0.47}
if(hist.sd.2[i]>=c.2) {sc.2<-0.47}
scale.2<-sc.2
scale.vec.2[i]<-sc.2
VaR.POT.vec.10.i.2<-VaR.POT.vec.save.2[i]*(10)^scale.2
VaR.POT.vec.10.2[i]<-VaR.POT.vec.10.i.2
VaR.POT.Mkn.vec.10.i.2<-VaR.POT.Mkn.vec.save.2[i]*(10)^scale.2
VaR.POT.Mkn.vec.10.2[i]<-VaR.POT.Mkn.vec.10.i.2
ES.POT.vec.10.2[i]<-VaR.POT.vec.10.i.2/(1-gamma.vec.save.2[i])+(sigma.vec.save.2[i]-
gamma.vec.save.2[i]*u.vec.save.2[i])/(1-gamma.vec.save.2[i])
ES.POT.Mkn.vec.10.2[i]<-VaR.POT.Mkn.vec.10.i.2/(1-M.k.n.vec.save.2[i])+(sigma.vec.save.2[i]-
M.k.n.vec.save.2[i]*u.vec.save.2[i])/(1-M.k.n.vec.save.2[i])
}

#Saving the values of ES and VaR for each of the 3 methods (and the ML and Mkn estimates):
#Square root of time rule:
VaR.POT.vec.10.SquareRoot.2<-VaR.POT.vec.10.2
ES.POT.vec.10.SquareRoot.2<-ES.POT.vec.10.2
VaR.POT.Mkn.vec.10.SquareRoot.2<-VaR.POT.Mkn.vec.10.2
ES.POT.Mkn.vec.10.SquareRoot.2<-ES.POT.Mkn.vec.10.2

#Alpha root rule:
VaR.POT.vec.10.AlphaRoot.2<-VaR.POT.vec.10.2
ES.POT.vec.10.AlphaRoot.2<-ES.POT.vec.10.2
VaR.POT.Mkn.vec.10.AlphaRoot.2<-VaR.POT.Mkn.vec.10.2
ES.POT.Mkn.vec.10.AlphaRoot.2<-ES.POT.Mkn.vec.10.2

#Empirically established scaling laws (McNeil):

```

```

VaR.POT.vec.10.McNeil.2<-VaR.POT.vec.10.2
ES.POT.vec.10.McNeil.2<-ES.POT.vec.10.2
VaR.POT.Mkn.vec.10.McNeil.2<-VaR.POT.Mkn.vec.10.2
ES.POT.Mkn.vec.10.McNeil.2<-ES.POT.Mkn.vec.10.2

#"Saving" of values to use for plots (Only VaR)

#ML+SQUARE ROOT:
VaR.POT.vec.10.SquareRoot.period1<-VaR.POT.vec.10.SquareRoot.2
VaR.POT.vec.10.SquareRoot.period3<-VaR.POT.vec.10.SquareRoot.2
VaR.POT.vec.10.SquareRoot.period5<-VaR.POT.vec.10.SquareRoot.2

#MOMENT+SQUARE ROOT:
VaR.POT.Mkn.vec.10.SquareRoot.period1<-VaR.POT.Mkn.vec.10.SquareRoot.2
VaR.POT.Mkn.vec.10.SquareRoot.period3<-VaR.POT.Mkn.vec.10.SquareRoot.2
VaR.POT.Mkn.vec.10.SquareRoot.period5<-VaR.POT.Mkn.vec.10.SquareRoot.2

#ML+ALPHA ROOT:
VaR.POT.vec.10.AlphaRoot.2.period1<-VaR.POT.vec.10.AlphaRoot.2
VaR.POT.vec.10.AlphaRoot.2.period3<-VaR.POT.vec.10.AlphaRoot.2
VaR.POT.vec.10.AlphaRoot.2.period5<-VaR.POT.vec.10.AlphaRoot.2

#MOMENT+ALPHA ROOT:
VaR.POT.Mkn.vec.10.AlphaRoot.period1<-VaR.POT.Mkn.vec.10.AlphaRoot.2
VaR.POT.Mkn.vec.10.AlphaRoot.period3<-VaR.POT.Mkn.vec.10.AlphaRoot.2
VaR.POT.Mkn.vec.10.AlphaRoot.period5<-VaR.POT.Mkn.vec.10.AlphaRoot.2

#ML+EMPIRICAL:
VaR.POT.vec.10.McNeil.2.period1<-VaR.POT.vec.10.McNeil.2
VaR.POT.vec.10.McNeil.2.period3<-VaR.POT.vec.10.McNeil.2
VaR.POT.vec.10.McNeil.2.period5<-VaR.POT.vec.10.McNeil.2

#MOMENT+EMPIRICAL
VaR.POT.Mkn.vec.10.McNeil.period1<-VaR.POT.Mkn.vec.10.McNeil.2
VaR.POT.Mkn.vec.10.McNeil.period3<-VaR.POT.Mkn.vec.10.McNeil.2
VaR.POT.Mkn.vec.10.McNeil.period5<-VaR.POT.Mkn.vec.10.McNeil.2

```

```

period1 corresponds to true.10.2[1:length(period1 forecast)]
period3 corresponds to true.10.2[1100:(1100+length(period3 forecast))]
period5 corresponds to true.10.2[3086:(3086+length(period5 forecast))]

#For the plots:
#ML+SQUARE ROOT and MOMENT+SQUARE ROOT:
par(mfrow=c(2,2))

#period1
plot(true.10.2[1:length(VaR.POT.vec.10.SquareRoot.period1)],col="lightblue",
main="Period 1",ty="l",ylim=c(min(true.10.2[1:length(VaR.POT.vec.10.SquareRoot.period1)],-
VaR.POT.vec.10.SquareRoot.period1,-VaR.POT.Mkn.vec.10.SquareRoot.period1
),max(true.10.2[1:length(VaR.POT.vec.10.SquareRoot.period1)],-
VaR.POT.vec.10.SquareRoot.period1,-VaR.POT.Mkn.vec.10.SquareRoot.period1)),
xlab="",ylab="")
lines(-VaR.POT.vec.10.SquareRoot.period1,col="red")
lines(-VaR.POT.Mkn.vec.10.SquareRoot.period1,col="green")

#period3
plot(true.10.2[1100:(1100+length(VaR.POT.vec.10.SquareRoot.period3))],col="lightblue",
main="Period
3",ty="l",ylim=c(min(true.10.2[1100:(1100+length(VaR.POT.vec.10.SquareRoot.period3))],-
VaR.POT.vec.10.SquareRoot.period3,-VaR.POT.Mkn.vec.10.SquareRoot.period3
),max(true.10.2[1100:(1100+length(VaR.POT.vec.10.SquareRoot.period3))],-
VaR.POT.vec.10.SquareRoot.period3,-VaR.POT.Mkn.vec.10.SquareRoot.period3)),
xlab="",ylab="")
lines(-VaR.POT.vec.10.SquareRoot.period3,col="red")
lines(-VaR.POT.Mkn.vec.10.SquareRoot.period3,col="green")

#period 5
plot(true.10.2[3086:(3086+length(VaR.POT.vec.10.SquareRoot.period5))],col="lightblue",
main="Period
5",ty="l",ylim=c(min(true.10.2[3086:(3086+length(VaR.POT.vec.10.SquareRoot.period5))],-
VaR.POT.vec.10.SquareRoot.period5,-VaR.POT.Mkn.vec.10.SquareRoot.period5
),max(true.10.2[3086:(3086+length(VaR.POT.vec.10.SquareRoot.period5))],-
VaR.POT.vec.10.SquareRoot.period5,-VaR.POT.Mkn.vec.10.SquareRoot.period5)),
xlab="",ylab="")
lines(-VaR.POT.vec.10.SquareRoot.period5,col="red")
lines(-VaR.POT.Mkn.vec.10.SquareRoot.period5,col="green")

plot(1:10,1:10,ty="n",xaxt="n",yaxt="n",xlab="",ylab="")
legend(x="topleft",col=c("lightblue","red","green"),lty=c(1,1,1),
legend=c("True 10 day returns","ML Estimate","Moment Estimate"))

```

```

#ML+ALPHA ROOT and MOMENT+ALPHA ROOT:

par(mfrow=c(2,2))

#period1

plot(true.10.2[1:length(VaR.POT.vec.10.AlphaRoot.2.period1)],col="lightblue",

main="Period 1",ty="l",ylim=c(min(true.10.2[1:length(VaR.POT.vec.10.AlphaRoot.2.period1)],-
VaR.POT.vec.10.AlphaRoot.2.period1,-VaR.POT.Mkn.vec.10.AlphaRoot.period1

),max(true.10.2[1:length(VaR.POT.vec.10.AlphaRoot.2.period1)],-
VaR.POT.vec.10.AlphaRoot.2.period1,-VaR.POT.Mkn.vec.10.AlphaRoot.period1)),

xlab="",ylab="")

lines(-VaR.POT.vec.10.AlphaRoot.2.period1,col="red")

lines(-VaR.POT.Mkn.vec.10.AlphaRoot.period1,col="green")

#period3

plot(true.10.2[1100:(1100+length(VaR.POT.vec.10.SquareRoot.period3))],col="lightblue",

main="Period
3",ty="l",ylim=c(min(true.10.2[1100:(1100+length(VaR.POT.vec.10.SquareRoot.period3))],-
VaR.POT.vec.10.AlphaRoot.2.period3,-VaR.POT.Mkn.vec.10.AlphaRoot.period3

),max(true.10.2[1100:(1100+length(VaR.POT.vec.10.SquareRoot.period3))],-
VaR.POT.vec.10.AlphaRoot.2.period3,-VaR.POT.Mkn.vec.10.AlphaRoot.period3)),

xlab="",ylab="")

lines(-VaR.POT.vec.10.AlphaRoot.2.period3,col="red")

lines(-VaR.POT.Mkn.vec.10.AlphaRoot.period3,col="green")

#period 5

plot(true.10.2[3086:(3086+length(VaR.POT.vec.10.SquareRoot.period5))],col="lightblue",

main="Period
5",ty="l",ylim=c(min(true.10.2[3086:(3086+length(VaR.POT.vec.10.AlphaRoot.2.period5))],-
VaR.POT.vec.10.AlphaRoot.2.period5,-VaR.POT.Mkn.vec.10.AlphaRoot.period5

),max(true.10.2[3086:(3086+length(VaR.POT.vec.10.SquareRoot.period5))],-
VaR.POT.vec.10.AlphaRoot.2.period5,-VaR.POT.Mkn.vec.10.AlphaRoot.period5)),

xlab="",ylab="")

lines(-VaR.POT.vec.10.AlphaRoot.2.period5,col="red")

lines(-VaR.POT.Mkn.vec.10.AlphaRoot.period5,col="green")

plot(1:10,1:10,ty="n",xaxt="n",yaxt="n",xlab="",ylab="")

legend(x="topleft",col=c("lightblue","red","green"),lty=c(1,1,1),

legend=c("True 10 day returns","ML Estimate","Moment Estimate"))

#ML+MCNEIL and MOMENT+MCNEIL:

par(mfrow=c(2,2))

#period1

```

```

plot(true.10.2[1:length(VaR.POT.vec.10.AlphaRoot.2.period1)],col="lightblue",
main="Period 1",ty="l",ylim=c(min(true.10.2[1:length(VaR.POT.vec.10.AlphaRoot.2.period1)],-
VaR.POT.vec.10.McNeil.2.period1,-VaR.POT.Mkn.vec.10.McNeil.period1
),max(true.10.2[1:length(VaR.POT.vec.10.AlphaRoot.2.period1)],-
VaR.POT.vec.10.McNeil.2.period1,-VaR.POT.Mkn.vec.10.McNeil.period1)),
xlab="",ylab="")
lines(-VaR.POT.vec.10.McNeil.2.period1,col="red")
lines(-VaR.POT.Mkn.vec.10.McNeil.period1,col="green")
#period3
plot(true.10.2[1100:(1100+length(VaR.POT.vec.10.SquareRoot.period3))],col="lightblue",
main="Period
3",ty="l",ylim=c(min(true.10.2[1100:(1100+length(VaR.POT.vec.10.SquareRoot.period3))],-
VaR.POT.vec.10.McNeil.2.period3,-VaR.POT.Mkn.vec.10.McNeil.period3
),max(true.10.2[1100:(1100+length(VaR.POT.vec.10.SquareRoot.period3))],-
VaR.POT.vec.10.McNeil.2.period3,-VaR.POT.Mkn.vec.10.McNeil.period3)),
xlab="",ylab="")
lines(-VaR.POT.vec.10.McNeil.2.period3,col="red")
lines(-VaR.POT.Mkn.vec.10.McNeil.period3,col="green")
#period 5
plot(true.10.2[3086:(3086+length(VaR.POT.vec.10.SquareRoot.period5))],col="lightblue",
main="Period
5",ty="l",ylim=c(min(true.10.2[3086:(3086+length(VaR.POT.vec.10.AlphaRoot.2.period5))],-
VaR.POT.vec.10.McNeil.2.period5,-VaR.POT.Mkn.vec.10.McNeil.period5
),max(true.10.2[3086:(3086+length(VaR.POT.vec.10.SquareRoot.period5))],-
VaR.POT.vec.10.McNeil.2.period5,-VaR.POT.Mkn.vec.10.McNeil.period5)),
xlab="",ylab="")
lines(-VaR.POT.vec.10.McNeil.2.period5,col="red")
lines(-VaR.POT.Mkn.vec.10.McNeil.period5,col="green")
plot(1:10,1:10,ty="n",xaxt="n",yaxt="n",xlab="",ylab="")
legend(x="topleft",col=c("lightblue","red","green"),lty=c(1,1,1),
legend=c("True 10 day returns","ML Estimate","Moment Estimate"))

```

Bibliography

Alexander, C. 2008(a). *Market Risk Analysis II: Practical Financial Econometrics*. 1st edition. England: John Wiley & Sons.

Alexander, C. 2008(b). *Market Risk Analysis IV: Value at Risk Models*. 1st edition. England: John Wiley & Sons.

Anon. 2011. *Deloitte: Market Risk*. [Online]. Available: <http://www.deloitte.com/assets/Dcom-SouthAfrica/Local%20Assets/Documents/7.%20Basel%20flyer%20-%20Market%20Risk.pdf> [2013, March 15].

Anon. 2012. *A Practical Introduction to GARCH Modeling*. [Online]. Available: <http://www.portfolioprobe.com/2012/07/06/a-practical-introduction-to-garch-modeling/> [2013, June 28].

Balkema, A. & de Haan, L. 1974. Residual Life Time at Great Age. *Annals of Probability*, 2: 792-804.

Beirlant, J., Goegebeur, Y., Segers, J. & Teugels, J. 2004. *Statistics of Extremes: Theory and Applications*. 1st edition. England: John Wiley & Sons.

Bensalah, Y. 2010. *Steps in Applying Extreme Value Theory to Finance: A Review*. [Online]. Available: www.bankofcanada.ca/wp-content/uploads/2010/01/wp00-20.pdf [2012, July 9].

Berning, T.L. 2013. *Quantification of estimation instability and its application to threshold selection in extremes*. In press.

Brooks, C., Clare, A.D., Dalle Molle, J.W. & Persaud, G. 2003. *A Comparison of Extreme Value Theory Approaches for Determining Value at Risk*. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0927539804000854> [2012, July 3].

Byström, H.N.E. 2004. Managing extreme risks in tranquil and volatile markets using conditional extreme value theory. *International Review of Financial Analysis*, 13:133-152.

Coles, S. 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag, London Ltd, London.

Danielsson, J. & de Vries, C.G. 2000. Value-at-Risk and Extreme Returns. *Annals d'Economie et de Statistique*, 60: 239-269.

de Haan, L. 1970. *On Regular Variation and its Applications to the Weak Convergence of Sample Extremes*. Mathematical Centre Tract, 32, Amsterdam, The Netherlands.

Dekkers, A.L.M., Einmahl, J.H.J. & de Haan, L. 1989. A Moment Estimator for the Index of an Extreme-value Distribution. *Annals of Statistics*, 17: 1833-1855.

Diebold, F. X., Shuermann, T. & Stroughair, J.D. 1998. *Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management*, Working Paper. [Online]. Available: www.ssc.upenn.edu/~fdiebold/papers/paper21/dss-f.pdf [2012, July 2].

Diebold, F.X., Hickman A. Inoue, A. & Shuermann, T. 1997. Converting 1-Day Volatility to h-Day Volatility: Scaling by Root-h is Worse than You Think. Wharton Financial Institutions Center, Working Paper 97-34. Published in condensed form as "Scale Models", *Risk*, 11, 104-107 (1998).

Engle, R. 2001. GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics. *Journal of Economic Perspectives*, 15(4): 157-168.

Fernandez, V. 2003. Extreme Value Theory and Value at Risk. *Revista de Análisis Económico*, 18(1): 57-85.

Fisher, R.A. & Tippett, L.H.C. 1928. On the Estimation of the Frequency Distributions of the Largest or Smallest Member of a Sample. *Proceedings of the Cambridge Philosophical Society*, 24: 180-190.

Ghalanos, A. 2013. *rugarch Version 1.2-1*. [Online]. Available: <http://cran.r-project.org/web/packages/rugarch/index.html> [2013, May 26].

Gilli, M. & Küllezi E. 2006. An Application of Extreme Value Theory for Measuring Financial Risk. *Computational Economics*, 27: 207-228.

Gnedenko, B.V. 1943. Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, 44: 423-453.

Gomes, M.I. & Martins, M.J. 2002. "Asymptotically Unbiased" Estimators of the Tail Index Based on External Estimation of the Second Order Parameter. *Extremes*, 5(1): 5-31.

Greenwood, J.A., Landwehr, J.M., Matalas, N.C. & Wallis, J.R. 1979. Probability Weighted Moments: Definition and Relation to Parameters of Several Distributions Expressible in Inverse Form. *Water Resources Research*, 15: 1049-1054.

Heffernan, J.E. 2009. *ismev Version 1.39*. Available: <http://cran.r-project.org/web/packages/ismev/index.html> [2013, June 19].

Hill, B.M. 1975. A Simple General Approach to Inference about the Tail of a Distribution. *Annals of Statistics*, 3: 1163-1174.

Hosking, J.R.M. 1985. Algorithm AS 215: Maximum Likelihood Estimation of the Parameters of the Generalized Extreme Value Distribution. *Applied Statistics*, 34: 301-310.

Hull, J.C. 2012. *Options, Futures and other Derivatives*. 8th Edition. USA: Pearson Education, Inc.

Ma, J., Nelson, C.R. & Startz, R. 2006. Spurious Inference in the GARCH(1,1) Model when it is Weakly Identified. *Studies in Nonlinear Dynamics and Econometrics*, 11(1): 1-27, Article 1.

McNeil, A. & Frey, R. 1998. *Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: an Extreme Value Approach*, preprint. ETH Zürich.

McNeil, A. & Frey, R. 2000. Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: an Extreme Value Approach. *Journal of Empirical Finance*, 7 (3-4): 271-300.

McNeil, A.J. 1999. *Extreme Value Theory for Risk Managers*. [Online]. Available: www.sfu.ca/~rjones/econ811/readings/McNeil%201999.pdf [2012, July 5].

Odening, M. & Hinrichs, J. 2002. Using Extreme Value Theory to Estimate Value-at-risk. *Agricultural Finance Review*, 63(1): 55-73.

Peng, Z., Li, S. & Pang, H. 2006. *Comparison of Extreme Value Theory and GARCH Models on Estimating and Predicting of Value-at-Risk*. [Online]. Available: <http://www.wise>.

xmu.edu.cn/seta2006/download/POSTER%20SECTION/%E9%BB%8E%E5%AE%9EPDF/xiamen_university_working_paper_20060114.pdf [2013, June 21].

Pickands, J. 1975. Statistical Inference Using Extreme Order statistics. *Annals of Statistics*, 3:119-131.

Ruppert, D. 2004. *Statistics and finance: An introduction*. New York: Springer-Verlag.

Weissmann, I. 1978. Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, 73: 812-815.

Zivot, E. 2008. *Practical Issues in the Analysis of Univariate GARCH Models*. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.186.312&rep=rep1&type=pdf> [2013, June 21].