Application of Extreme Value Theory to the Calculation of Value-at-Risk

Anthony Seymour

Department of Mathematics and Applied Mathematics

Faculty of Science

University of Cape Town

September 2001

Supervisor: Dr D Polakow

Acknowledgements

I would like to express my sincere appreciation and gratitude toward the following:

Dr Daniel Polakow for his interest, guidance and encouragement throughout the study.

Dr Peter Ouwehand for his helpful discussions and general assistance.

My fellow Mathematics of Finance students, especially Steven Becker, Dug Binyana, Fernando Durrel, and Clare Johnson.

My family, and Sophie Rees-Jones for never-ending support.

Investec Ltd for financial assistance.

Summary

An investigation was carried out to examine the application of extreme value theory (EVT) to the calculation of value-at-risk (VaR). The main aim of the study was to test the applicability of published EVT-based VaR calculation methods to the South African market. Two methods were tested on a hypothetical portfolio of South African stocks, using the standard backtesting technique. Both methods can be viewed as simple extensions of two of the most popular standard VaR methodologies, *viz.* historical simulation and J.P. Morgan's RiskMetrics™, which makes use of GARCH volatility modelling. The RiskMetrics™-based EVT method was found to be superior to historical simulation (popular in South Africa) as well as to the EVT extension of historical simulation, thus highlighting the importance of incorporating volatility updating into the VaR calculation.

Contents

Summary

1 Introduction 1

2 Overview of Extreme Value Theory 8

2.1 Block Maxima Models 8

2.2 The Peaks over Threshold Method 11

2.3 Overview of GARCH volatility models 12

3 Extreme Value Theory and Value-at-Risk 15

3.1.1 Danielsson and De Vries' Tail Estimator 16

3.1.2 Derivation of Danielsson and De Vries' Tail Estimator 16

3.1.3 Results Obtained by Danielsson and De Vries 17

3.2.1 McNeil and Frey's Tail Estimator 19

3.2.1 Derivation of McNeil and Frey's Tail Estimator 19

3.2.3 Implementation and Results 22

4 Application of EVT-Based VaR Calculation Methods to the South African Market 25

4.1 Selection of Stocks 25

4.2 Historical Simulation 29

4.3 Danielsson and De Vries' Tail Estimator 30

4.4 McNeil and Frey's Tail Estimator 37

4.5 Asymmetric GARCH Models 49

5. Conclusion 53

References 54

Appendix A 56

Appendix B 57

Appendix C 59

Appendix D 63

Chapter 1

Introduction

Recent large losses incurred by certain organisations in the course of their trading activities have highlighted the need for effective risk management. A well known example is the case of Orange County, which in December 1994 announced that its investment pool had suffered a loss of $1.6 billion. Philippe Jorion describes how the loss was the result of the unsupervised investment activity of Bob Citron, the County Treasurer.[1] Similarly, the demise of Barings Bank has been attributed to the actions of a single individual. It is clear that in both these cases, risk management was poor or non-existent.

In order to monitor and manage the risks inherent in the general trading activities of an organisation, senior management require some sort of quantitative measure of risk. A popular approach to quantifying the risk associated with a portfolio of financial assets is simply to use the variability of the historical returns of the portfolio as an indication of risk. A portfolio that in the past has delivered highly variable returns is generally perceived to be more risky than an asset that affords more consistent returns. Statistically, this means using the *standard deviation* of the portfolio returns as a measure of risk. This perception is, however, most suited to situations in which the *relative* riskiness of different portfolios is of interest. An alternative indication of risk involves calculating the sensitivity of the portfolio to specific market variables. Established measures, connected to derivatives, include *delta*, which is a measure of the sensitivity of the price of an option to changes in the value of the underlying, and *duration,* a measure of the sensitivity of the price of a bond to changes in interest rate. Measures such as these are, however, technical in nature, and are of most use to traders who are responsible for managing the various components of the financial institution's portfolio. They are thus of limited use to non-specialists and senior management who simply want a concise summary of the risks to which the company is exposed. The most widely used measure for this purpose at the present time is value-at-risk (VaR). VaR was developed to provide a single number that could encapsulate information about the risk in a portfolio.[2]

A formal definition of VaR can be given as "an estimate of the level of loss on a portfolio which is expected to be equaled or exceeded with a given, small probability".[2] An alternative definition can be given in terms of the statement one wishes to make when calculating VaR: "We are π percent certain that we will not lose more than V dollars in the next N days".[3] Here, V is the VaR of the portfolio. This definition is perhaps better in that it demonstrates that VaR is a function of π, the confidence level, and N, the time horizon.

In order to calculate VaR, it is necessary to cast the preceding definitions into standard mathematical notation. If X is defined to be the loss on a portfolio, and π the confidence level, then VaR is simply defined by

$$\mathbb{P}(X \leq VaR) = \pi$$

This is equivalent to

$$F_X(VaR) = \pi$$

where $F_X(x) = \mathbb{P}(X \leq x)$ is the probability distribution function of the random variable X. It is thus clear that if the probability distribution function of X is known, then VaR is given by

$$F_X^{-1}(\pi) = VaR$$

where $F_X^{-1}(\bullet)$ denotes the inverse of $F_X(\bullet)$. The problem of calculating VaR therefore translates to the specification of the probability distribution function $F_X(\bullet)$. All methods of VaR analysis are simply alternative ways of determining the profit and loss distribution function of a portfolio.

There are three main methods of VaR calculation. These are: parametric, historical simulation, and Monte Carlo.

Parametric

The parametric method is used by RiskMetrics™ in its most basic form. It involves specifying the type of distribution for the profit and loss variables, and using analytical techniques to calculate the volatility (standard deviation σ) of the portfolio. Generally, the profits/losses on a portfolio are assumed to be normally distributed with a mean of zero, thus allowing the probability distribution function to be fully specified once the current volatility is known. The value of σ is iteratively recalculated as a one-step ahead predicted value obtained via a GARCH-type model (GARCH models are discussed further in chapter 2). In particular, for a normally distributed random variable X, the value of x such that

$$\mathrm{P}(X \le x) = 0.95$$

is given by $1.65\sigma_X$. Thus in the simple RiskMetrics™ framework, the 95% VaR is given by 1.65 × (volatility of the portfolio). Similarly, the 99% VaR is given by 2.33 × (volatility of the portfolio).[4]

The most common criticism against this type of approach is that the assumption that the profits/losses on a portfolio are normally distributed is unrealistic. It is known that financial return series exhibit *leptokurtosis*, or "heavier tails" than a normal distribution.[5] In essence, this knowledge means that any VaR calculation technique based on a normal distribution function will tend to give VaR estimates that are too low.

The advantage of the RiskMetrics™ methodology is that the current volatility background is captured via a GARCH model, allowing VaR estimates to take into account changing volatility.

Historical Simulation

The simplest method of obtaining a profit and loss distribution of a portfolio is via a method called *historical simulation*. The method requires a database of returns for the stocks comprising the portfolio, and uses these returns to generate a set of simulated returns for the portfolio. We consider a database consisting of N days returns for K stocks in a portfolio. The return for stock k on day i is defined as

$$r_i^k = \frac{S_i^k - S_{i-1}^k}{S_{i-1}^k}$$

where S_i^k is the price of stock k on day i. The return on the portfolio on day i is given by

$$R_i = \sum_{k=1}^{K} r_i^k \omega_k$$

where $(\omega_1,\ldots,\omega_K)$ is a vector of weights (by value) of stocks in the portfolio $\left(\sum_{k=1}^{K} \omega_k = 1\right)$.

The calculation to obtain returns for the portfolio for each of the N days can be represented in matrix notation as

$$\begin{pmatrix} r_1^1 & r_1^2 & \cdots & r_1^K \\ r_2^1 & r_2^2 & \cdots & r_2^K \\ \vdots & \vdots & \ddots & \vdots \\ r_N^1 & r_N^2 & \cdots & r_N^K \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_K \end{pmatrix} = \begin{pmatrix} R_1 \\ R_2 \\ \vdots \\ R_N \end{pmatrix}$$

where r_i^k denotes the return of stock k on day i, as defined above, and R_i is the return on the portfolio on day i. The returns vector can be used to obtain an empirical distribution function for the losses on the portfolio by converting the returns to losses (simply via a change of sign), and sorting the losses into an ordered set

$$X_1 \leq \ldots \leq X_N$$

Then the empirical probability distribution for the losses on the portfolio is defined by

$$\mathrm{P}(X \leq x) = \begin{cases} 0 & x < X_1 \\ i/N & X_i \leq x \leq X_{i+1} \\ 1 & X_N \leq x \end{cases}$$

An example of how such an empirical distribution function may appear is shown in Figure 1.1. The plot was created by generating 20 normally distributed iid random variables and applying the above definition. The "staircase" appearance of the plot reflects the discrete nature of the data used.

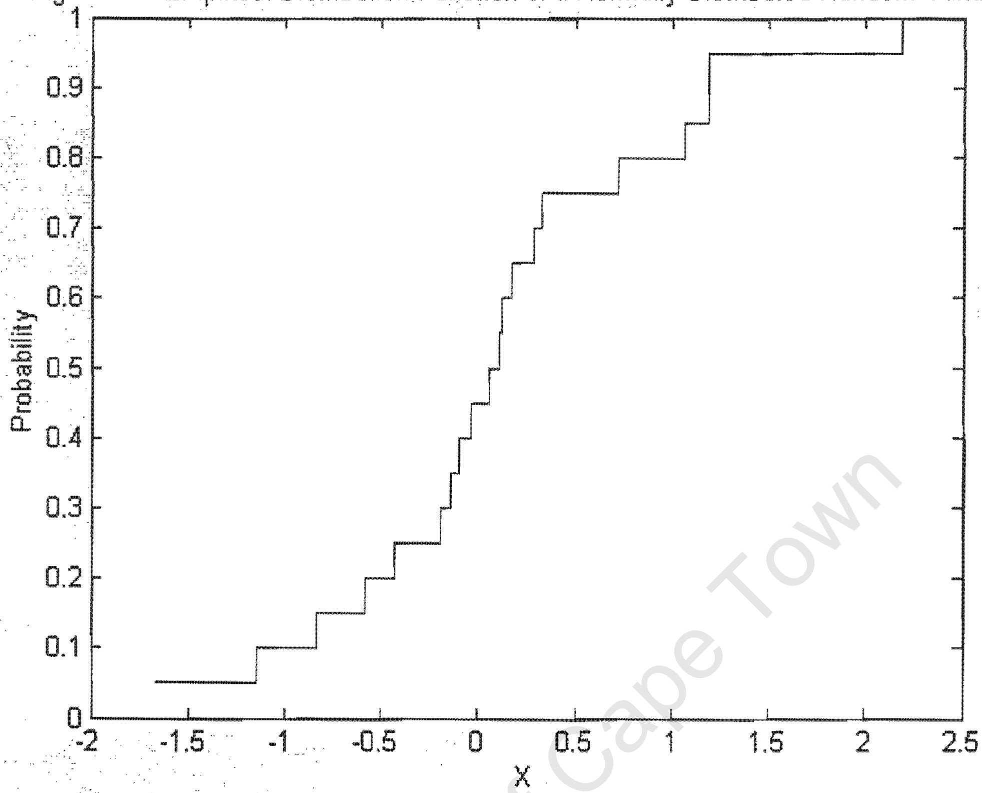
For a data set comprised of N days, and X representing the loss on a portfolio, VaR is given by X_i where

$$i = N \times \pi$$

Thus for a risk confidence level of $\pi = 0.95$, and a data set of 100 observations, VaR would be given as the 95th loss in the ordered set.

The main disadvantage to this approach is that it is impossible to obtain a VaR estimate for an out-of-sample probability. In other words, the smallest probability that can be obtained such that $\mathrm{P}(X \leq VaR)$ is $1/N$. Thus for a sample size of 100, it would not be possible to estimate VaR at a confidence level higher than 99%. The advantage of the historical simulation approach is that it accurately reflects the historical probability distribution of the market variables. However, since relatively few data points of a finite real-life data set enter the tail of the distribution, VaR estimates for extreme probabilities obtained from these observations are likely to be inaccurate.

Figure 1.1: Empirical Distribution Function of a Normally Distributed Random Variable



Monte Carlo

The third common method of VaR calculation involves generating the profit and loss probability distribution function via Monte Carlo simulation.[6] This method is similar to that of historical simulation, except that the returns for the stocks comprising the portfolio are drawn randomly from a pre-specified distribution (usually normal), as opposed to actual historical data. The method suffers from the same drawbacks as historical simulation, as well as possible mis-specification of the distribution of the underlying stocks. In addition, as in any Monte Carlo simulation, the method is computationally intensive and therefore tends to be slow.

It is clear that VaR has an important role to play in risk management, but its use has been criticized, mainly owing to the deficiencies mentioned here. Recently, an established branch of statistics called extreme value theory (EVT) has been employed to address some of these shortcomings. EVT is concerned with extremes, or rare events, and is therefore tailor-made for VaR estimation. The aim of the study was to identify published

EVT-based VaR calculation methods that had been successfully applied in foreign markets, and to test their applicability in estimating VaR in a South African context. The following chapter provides a brief overview of EVT.

Chapter 2

Overview of Extreme Value Theory

One of the methods outlined in Chapter 1 for VaR estimation involves assuming a normal distribution for portfolio returns, which can lead to a serious underestimation of VaR, owing to the "heavy tailed" nature of financial return series. A possible alternative to this method is to fit a "fat tailed" distribution to all the data. The drawback of this type of strategy is that the parameters of the distribution are constructed to fit the central, common observations best, and the estimated distribution is therefore ill-suited to the extreme observations with which VaR estimation is concerned. The problem facing anyone attempting a VaR analysis is one of estimating the probabilities of extreme or rare events with limited data. This type of problem is not unique to risk management, occurring in other fields such as hydrology where engineers have had to determine how high sea walls should be to contain flood probabilities within reasonable limits.[8] The branch of statistics that has arisen out of a study of these types of problems is extreme value theory (EVT). In essence, EVT allows one to determine the nature of the tail of a distribution without the need for strong assumptions concerning the distribution from which the data are drawn.

EVT provides two main approaches to the analysis of extremes. The older of the two involves the use of *block maxima models*. These are models for the largest observations collected from large samples of identically distributed observations. For example, if daily profits and losses on a portfolio are recorded and assumed iid, the block maxima method provides a model that may be appropriate for the quarterly maximum of such values.[9]

The second approach is known as the *peaks-over-threshold (POT)* method, and is for all large observations that exceed a high threshold. The POT models are generally considered to be the most useful for practical applications, owing to their more efficient use of the limited data on extreme values.[9]

2.1 Block Maxima Models

The block maxima models concern the maxima of a set of observations. The key result for this type of analysis is the Fisher-Tippet theorem, which gives weak convergence results for centred and normalised maxima. The result is similar in spirit to the central limit theorem, which gives convergence results for sequences of normalised *sums* of random variables. The Fisher-Tippet theorem is given formally as:

Theorem 2.1.1 (Fisher Tippet theorem, limit laws for maxima)[10]
Let (X_n) be a sequence of iid random variables, and let $M_n = max(X_1, ...X_n)$. If there exist norming constants $c_n > 0$, $d_n \in \mathbb{R}$ and some non-degenerate distribution function H such that

$$\frac{M_n - d_n}{c_n} \xrightarrow{d} H$$

then H belongs to one of the following three distribution function families:

$$\text{Fréchet}: \qquad \Phi_\alpha(x) = \begin{cases} 0 & x \le 0 \\ \exp\{-x^{-\alpha}\} & x > 0 \end{cases}$$

$$\text{Weibull}: \qquad \Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^\alpha\} & x \le 0 \\ 1 & x > 0 \end{cases}$$

$$\text{Gumbel}: \qquad \Lambda(x) = \exp\{-e^{-x}\} \quad x \in \mathbb{R}$$

The Fisher-Tippet theorem thus states that the only possible limit laws for maxima M_n, when properly normalised and centred, are Φ_α, Ψ_α, and Λ. These distributions are known as the *standard extreme value distributions*. They can be conveniently represented by a one parameter representation, known as the Jenkinson-von Mises representation:

Definition 2.1.1 (Jenkinson-von Mises representation of the extreme value distributions: the generalised extreme value distribution (GEV))[11]

Define the distribution function H_ξ by

$$H_\xi(x) = \begin{cases} \exp\left\{-(1+\xi x)^{-1/\xi}\right\} & \xi \neq 0 \\ \exp\left\{-\exp\left\{-x\right\}\right\} & \xi = 0 \end{cases}$$

where $1+\xi x > 0$.

$\xi = 1/\alpha > 0$ *corresponds to the Fréchet distribution* Φ_α,
$\xi = 0$ *corresponds to the Gumbel distribution* Λ,
$\xi = -1/\alpha < 0$ *corresponds to the Weibull distribution* Ψ_α.

An important concept in extreme value theory is that of the *maximum domain of attraction* of an extreme value distribution. This groups distribution functions for which the normalised maxima M_n have the same limit distribution into a common class, and is defined as follows:

Definition 2.1.2 (Maximum domain of attraction)[12]

We say that the distribution function F of the random variable X belongs to the maximum domain of attraction of the extreme value distribution *H if there exist constants* $c_n > 0$, $d_n \in \mathbb{R}$ *such that*

$$\frac{M_n - d_n}{c_n} \xrightarrow{d} H$$

We write $F \in \text{MDA}(H)$.

For each extreme value distribution it is possible to characterise its maximum domain of attraction. For example, it can be shown that

$$F \in \text{MDA}(\Phi_\alpha) \Leftrightarrow \lim_{x\to\infty} \frac{\overline{F}(xt)}{\overline{F}(x)} = t^{-\alpha}, \quad t > 0$$

$(\overline{F}(x) = 1 - F(x))$

In other words, the maximum domain of attraction of the Fréchet distribution Φ_α consists of distribution functions whose right tail is regularly varying with index $-\alpha$.

This class of distribution functions contains very heavy tailed distributions, and thus may be appropriate distributions for modelling financial return series.[13] Another important point is that these distributions are Pareto-like in the sense that their right tails are of the form $\bar{F}(x) \sim Kx^{-a}, \quad x \to \infty$.

2.2 The Peaks Over Threshold Method

Central to the POT method are the *excess distribution function* and the *generalised Pareto distribution* (*GPD*). The excess distribution function represents the probability that a loss exceeds a high threshold by a certain amount, given the information that it exceeds the threshold. A formal definition is:

Definition 2.2.1 (Excess distribution function) [14]

Let X be a random variable with distribution function F and right end point x_F $(x_F = \sup\{x \in \mathbb{R} : F(x) < 1\})$. For a fixed $u < x_F$,

$$F_u(x) = \mathbb{P}(X - u \leq x \mid X > u), \quad x \geq 0,$$

is the excess distribution function of the random variable X over the threshold *u*.

The GPD is a two-parameter distribution with distribution function[9]

$$G_{\xi,\beta}(x) = \begin{cases} 1-(1+\xi x/\beta)^{-1/\xi} & \xi \neq 0 \\ 1-\exp\{-x/\beta\} & \xi = 0 \end{cases}$$

where $\beta > 0$, and where $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\beta/\xi$ when $\xi < 0$.

The GPD is heavy tailed when $\xi > 0$, hence this case is the most relevant for risk management purposes.

The key result in the POT method is that the GPD appears as the limit distribution of the excess distribution function over high thresholds. This result is given formally by the following theorem:

Theorem 2.2.1 (Property of GPD) [15]

For every $\xi \in \mathbb{R}$, $F \in MDA(H_\xi)$ *if and only if*

$$\lim_{u \uparrow x_F} \sup_{0<x<x_F-u} \left| F_u(x) - G_{\xi,\beta(u)}(x) \right| = 0$$

for some positive function β.

This result suggests the following approximation:

$$\bar{F}_u(y) \approx \bar{G}_{\xi,\beta(u)}(y)$$

Using this approximation, one is able to derive a tail estimator, the details of which are discussed later.

One of the challenges in the application of EVT is the choice of threshold, or equivalently, the start of the tail of the distribution. It is clear that the GPD approximation of the excess distribution function is dependent on the choice of threshold, but there are no strict guidelines on how to select it. The problem is basically one of a trade-off between bias and variance.[16] The choice of threshold should be large enough to satisfy the conditions of its application (which may lead to a high variance in the tail estimates), while at the same time leaving sufficient observations for the estimation (which may give rise to a bias). There are graphical techniques to assist in the selection of the threshold,[14] and a recent paper by Danielsson and De Vries describes a bootstrap procedure to calculate an optimal threshold.[17] However, this aspect of EVT still contains subjective elements and this fact should be kept in mind.

2.3 Overview of GARCH volatility models

Owing to the crucial role of GARCH volatility models in one of the EVT-based VaR calculation methods that was investigated, an overview of these models is presented here. An important feature of financial return series is that they exhibit time-varying variance, or *heteroskedasticity*. In other words, the volatility of such series is not constant. In particular, they exhibit volatility clustering, in which large changes tend to follow large changes, and small change tend to follow small changes.[18] A model of the conditional variance to account for this type of behaviour was proposed by Engle in 1982, called *autoregressive conditional heteroskedasticity* (*ARCH*).[19] In 1986, Bollerslev published a modification of ARCH, called *generalised autoregressive conditional heteroskedasticity* (*GARCH*).[20] Bollerslev designed GARCH to offer a more parsimonious model (i.e. using fewer parameters) that lessens the computational burden.[18] The following section gives a technical description of GARCH, and how it relates to standard time series analysis.

A process that is often used to model the dynamics of financial return series is the *autoregressive process*. A general autoregressive process of order p, denoted AR(p),[21] satisfies

$$Y_t = C + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t$$

where C, ϕ_i, i=1...p are constants. The p indicates that the variable Y_t is a function of the previous p observations of the process (autoregressive terms). Also included in the process is a noise term. In particular, $\{\varepsilon_t\}$ could be a strict white noise sequence satisfying

$$\begin{aligned} &\mathbb{E}(\varepsilon_t) = 0 \\ &\mathbb{E}(\varepsilon_t^2) = \sigma^2 \\ &\mathbb{E}(\varepsilon_t \varepsilon_\tau) = 0 \quad t \neq \tau \end{aligned}$$

It is also possible to include what are called *moving average terms* to give an autoregressive moving average process, denoted by ARMA(p,q):[22]

$$Y_t = C + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

Assume, without loss of generality, that a financial return series follows an autoregressive process of order 1, i.e.

$$Y_t = \phi Y_{t-1} + \varepsilon_t$$

The above definition for the noise process is of a process whose variance is constant with respect to time. However, as mentioned previously, return series do not exhibit constant variance. To take into account the effect of time-varying variance, the noise term can be expressed as the product of two terms, affording the following AR(1) process:

$$Y_t = \phi Y_{t-1} + \sigma_t Z_t$$

where Z_t is independent, identically distributed with zero mean and unit variance, and σ_t^2 evolves according to

$$\sigma_t^2 = \kappa + \sum_{i=1}^{r} \alpha_i \sigma_{t-i}^2 + \sum_{k=1}^{m} \beta_k \varepsilon_{t-k}^2$$

and κ, α_i, β_k $(i = 1...r,\ k = 1...m)$ are constants. This is Bollerslev's *generalised autoregressive conditional heteroskedasticity* model, denoted by

$$\varepsilon_t \sim GARCH(r, m)$$

The term "autoregressive" indicates that there is a dependence on previous estimates of the variance, and conditional implies a dependence on the observations of the immediate past. By using the GARCH framework, the noise term in an ARMA process can be viewed as a random variable drawn from a particular distribution whose variance is changing through time.

The details concerning the estimation of the GARCH parameters, as well as the implementation of the model will be discussed later in the report.

Chapter 3

Extreme Value Theory and Value-at-Risk

The topic of the estimation of tails of profit and loss distribution functions via EVT has been the subject of extensive recent research. Kevin Dowd[8] attributes the first application of EVT to the problem of VaR estimation to Francois Longin,[23] who uses the block maxima approach described in Chapter 2. Subsequent work has focussed on similar methods, or on approaches based on the excesses over a high threshold method.[24] The important distinction to be made between the approaches is based not on the particular EVT technique used, but the type of profit and loss distribution that is estimated. One may focus on a distribution that ignores the stochastic volatility structure (as in the historical simulation method), or one may concentrate on a distribution that reflects the current volatility background. The former type of distribution is known as the marginal or stationary distribution of the process; the latter is known as the conditional return distribution where the conditioning is on the current volatility.[24d] At the present time, and to the best of our knowledge, only the method presented by McNeil and Frey incorporates volatility updating into an EVT-based VaR estimation technique.[24d] All other methods focus on the estimation of the stationary distribution, a method which McNeil and Frey refer to as "static EVT".

The aim of the study is to test the applicability of EVT-based VaR estimation methods to the South African market. After a survey of the published methods, two approaches were selected for testing, *viz*. the methods proposed by Danielsson and De Vries,[24b] and by McNeil and Frey.[24d] These methods were chosen because they demonstrate complementary approaches to incorporating EVT into VaR estimation (i.e. marginal or unconditional vs. conditional distribution). Furthermore, both methods can be viewed as extensions of established methodologies, thus providing scope for the comparison of "traditional" VaR estimation techniques with an EVT based approach.

3.1.1 Danielsson and De Vries' Tail Estimator

In a nutshell, the method of Danielsson and De Vries can be viewed as an extension of historical simulation. The approach essentially involves obtaining an empirical distribution function via historical simulation, and using EVT to fit a smooth curve through the tail of this distribution. It thus appears to be an attractive method for implementation in South Africa, as historical simulation is already in use by prominent organisations such as Investec and PSG Investment Bank, and the EVT estimator can be applied without much extra effort.

A noteworthy aspect of Danielsson and De Vries' approach is that it does not take into account the stochastic nature of volatility. Their reasons for adopting this kind of method are based on empirical observations of extreme returns. In particular, they state that extremes (or spikes) exhibit no volatility clustering, and no strong correlation. The following quote summarises their defence of their approach: "Extreme returns occur infrequently, and in addition do not appear to be related to a particular level of volatility. Therefore, an unconditional approach is better suited for VaR estimation than conditional volatility forecasts".[24b]

3.1.2 Derivation of Danielsson and De Vries' Tail Estimator

The derivation of Danielsson and De Vries' tail estimator given here follows the approach adopted by Embrechts *et al.*,[25] to ensure consistency throughout the report. The profit and loss distribution F is assumed to be in the maximum domain of attraction of the Fréchet distribution, as this class contains all heavy tailed distributions (see Chapter 2). Thus, F has a Pareto-like tail, i.e.

$$\bar{F}(x) = Cx^{-\alpha}$$

for x above a certain threshold u. Maximum likelihood estimation, assuming k iid tail values above u, can be applied to obtain estimates for C and α, *viz.*

$$C = \frac{k}{n} X_{k,n}^{\alpha}; \quad \frac{1}{\alpha_{k,n}} = \frac{1}{k}\sum_{j=1}^{k} \ln X_{j,n} - \ln X_{k,n}$$

where k is the number of observations above the threshold $u = X_{k,n}$ and

$$X_{k,n} \leq \ldots \leq X_{1,n},$$

the observations above the threshold, in ascending order. The value n is the total number of observations in the sample. The estimator for α is known as the *Hill estimator*.

Combining the estimators for C and α affords a tail estimate for the distribution function of losses as

$$1 - F(x) = \bar{F}(x) = \frac{k}{n}\left(\frac{x}{X_{k,n}}\right)^{-\alpha_{k,n}}, \quad x > X_{k,n}$$

In calculating VaR, one is interested in a number x_p such that

$$F(x_p) = \mathbb{P}(X \leq x_p) = p$$

Inverting the tail estimator affords an estimate of x_p (known as the p^{th} quantile):

$$x_p = \left(\frac{np}{k}\right)^{-\frac{1}{\alpha_{k,n}}} X_{k,n}$$

3.1.3 Results obtained by Danielsson and De Vries

The VaR estimation procedure of Danielsson and Devries is implemented simply by selecting the number of observations that comprise the tail (i.e. k), and applying the quantile estimation formula above. Danielsson and De Vries tested their estimator via a technique known as *backtesting*. Essentially, this involves calculating VaR a number of times, and comparing each VaR estimate to the realised return the following day. If, for example, a 95% confidence level VaR estimate was violated (or exceeded) on approximately 5% of occasions, then the VaR methodology may be believed to be sound. The data set for each VaR estimation was set to 1500 trading days, and 6 US stocks were randomly selected in addition to the J.P. Morgan bank stock as the basis for portfolio analysis. The strategy for the backtest was to generate 500 random portfolio weights, which represent 500 different portfolios containing the 7 stocks. For each of these portfolios, VaR was calculated by applying the portfolio weights to 1500 days of returns for the individual stocks to obtain 1500 days of returns for the portfolio, as in the method of historical simulation, and using the EVT-derived estimator. The VaR estimate was compared to the realised portfolio return on the following day, and this procedure was repeated for a consecutive set of 1000 days. In summary, VaR was calculated 1000 times for each of 500 different portfolios, and the number of violations or exceedances of the VaR estimate counted. The total number of trading days required for this backtest was thus 2500.

The results of Danielsson and De Vries' backtest are shown in Table 3.1. The number of violations of the VaR estimates are given as averages of the violations for the 500 portfolios, with standard errors in parenthesis. In addition, the average number of violations of VaR estimates at various confidence levels are presented. These averages can be compared with the expected number of violations for each confidence level. Table 3.1 also displays the results of backtests of the basic RiskMetrics™ methodology, and historical simulation.

Table 3.1 Violations of Danielsson & De Vries' VaR Estimator

		Method		
Confidence Level	**Expected**	**RiskMetrics**	**Historical Simulation**	**EVT**
0.95000	50.00	52.45 (7.39)	43.24 (10.75)	43.14 (11.10)
0.97500	25.00	30.26 (4.41)	20.50 (7.22)	20.84 (7.35)
0.99000	10.00	16.28 (3.13)	7.66 (3.90)	8.19 (3.86)
0.99500	5.00	10.65 (2.73)	3.69 (2.39)	4.23 (2.55)
0.99725	2.50	7.29 (2.27)	1.90 (1.57)	2.35 (1.72)
0.99900	1.00	4.85 (2.06)	0.95 (1.03)	1.06 (1.13)
0.99950	0.50	3.55 (1.81)	0.75 (0.89)	0.59 (0.82)
0.99725	0.25	2.72 (1.66)	0.75 (0.89)	0.33 (0.62)
0.99990	0.10	2.00 (1.45)	0.75 (0.89)	0.12 (0.35)
0.99995	0.05	1.58 (1.29)	0.75 (0.89)	0.06 (0.23)

RiskMetrics™ is seen to have provided the most accurate VaR estimates at the lowest confidence level, but consistently under-predicted VaR at subsequent levels. Historical simulation performed well until its probability limit (1/1500), and the EVT-based estimator is impressive in its agreement with the expected number of violations, especially at the higher confidence levels. Based on these results, one may conclude that for the data set employed by Danielsson and De Vries, their estimator seems appropriate.

3.2.1 McNeil and Frey's Tail Estimator

The distinction between the method of McNeil and Frey, and that of Danielsson and De Vries, is the type of distribution function that is estimated. McNeil and Frey disagree with Danielsson and De Vries' rationale that accounting for stochastic volatility in the analysis of extreme returns is unnecessary, and propose a method which incorporates GARCH volatility modelling to estimate a *conditional* distribution. In particular, if X_t represents the loss on a financial asset price at time t, the distribution of interest is

$$F_{X_{t+1}|\mathcal{G}_t}(x) = \mathbb{P}(X_{t+1} \leq x | \mathcal{G}_t)$$

which represents the predictive distribution of the loss over the next day, given knowledge $\mathcal{G}_t$ up to and including day t. VaR estimation thus involves estimating quantiles in the tails of distributions such as these.

3.2.2 Derivation of McNeil and Frey's Tail Estimator

The key to McNeil and Frey's approach is the specification of the dynamics of $(X_t, t \in \mathbb{Z})$, the time series representing daily observations of the loss on a financial asset price. It is assumed that the dynamics of X_t are given by

$$X_t = u_t + \sigma_t Z_t$$

where $u_t = \phi X_{t-1}$ (ϕ constant), and Z_t is a strict white noise process (independent, identically distributed) with zero mean and unit variance. This model has an expectation structure that is AR(1), with changing variance. The distribution function of interest is

$$\mathbb{P}(X_{t+1} \leq x | \mathcal{G}_t)$$

Substituting the expression for X_t above affords

$$\begin{aligned}
\mathbb{P}(u_{t+1} + \sigma_{t+1} Z_{t+1} \leq x | \mathcal{G}_t) &= \mathbb{P}\left(Z_{t+1} \leq \frac{x - u_{t+1}}{\sigma_{t+1}} \middle| \mathcal{G}_t \right) \\
&= \mathbb{P}\left(Z_{t+1} \leq \frac{x - u_{t+1}}{\sigma_{t+1}} \right) \\
&= F_Z\left(\frac{x - u_{t+1}}{\sigma_{t+1}} \right)
\end{aligned}$$

where $F_Z(z)$ denotes the marginal distribution function of Z_t.

We seek the q^{th} quantile of the predictive distribution of X_t, i.e. x_q such that

$$\mathbb{P}(u_{t+1} + \sigma_{t+1} Z_{t+1} \leq x_q | \mathcal{G}_t) = q$$

As shown above, this is equivalent to

$$\mathbb{P}\left(Z_{t+1} \leq \frac{x_q - u_{t+1}}{\sigma_{t+1}} \right) = q$$

Thus if z_q represents the q^{th} quantile of $F_Z(z)$, i.e. $F_Z(z_q) = q$, then

$$\frac{x_q - u_{t+1}}{\sigma_{t+1}} = z_q \quad \Leftrightarrow \quad x_q = u_{t+1} + \sigma_{t+1} z_q \; .$$

We thus have an expression relating the q^{th} quantile of the conditional distribution of the loss series to the q^{th} quantile of the white noise process. McNeil and Frey propose estimating the conditional mean u_{t+1} via an AR(1) process, and σ_{t+1} via a GARCH(1,1) model. An important part of their approach involves using EVT to model the tail of the distribution of the white noise process to obtain an estimate of the underlying quantile z_q. This objective is in contrast to the method of Danielsson and De Vries who use EVT to model the tail of the returns distribution itself.

McNeil and Frey adopt the peaks-over-threshold (POT) EVT technique to model the tail of $F_Z(z)$. Recall from Chapter 2 that the key result from this branch of EVT is that the distribution function of excesses over a high threshold is approximately modelled by a generalised Pareto distribution, i.e.

$$F_u(y) \approx G_{\xi,\beta(u)}(y)$$

Note that

$$\begin{aligned} F_u(y) &= \mathbb{P}\left(Z-u \le y \middle| Z>u\right) \\ &= \frac{\mathbb{P}\left(\{Z-y \le y\} \cap \{Z>u\}\right)}{\mathbb{P}(Z>u)} \\ &= \frac{\mathbb{P}(u<Z \le u+y)}{\mathbb{P}(Z>u)} \\ &= \frac{F(u+y)-F(u)}{1-F(u)} \end{aligned}$$

To employ the POT method, it is assumed that the following equality holds:

$$\frac{F(u+y)-F(u)}{1-F(u)} = G_{\xi,\beta}(y)$$

Let $z = u + y$. Then

$$\frac{F(z)-F(u)}{1-F(u)} = G_{\xi,\beta}(z-u)$$

$$\Leftrightarrow F(z) = \left(1-F(u)\right)G_{\xi,\beta}(z-u)+F(u) \qquad \textbf{(3.2.1)}$$

for $z > u$.

Equation **3.2.1** is used to obtain a tail estimator for F, once estimates of ξ, β and $F(u)$ have been calculated. If N_u out of a total of n data points exceed the threshold u, the GPD is fitted to the N_u excesses via maximum likelihood estimation to afford estimates $\hat{\xi}$ and $\hat{\beta}$. An estimate of $F(u)$ is obtained using the empirical estimator

$$\hat{F}(u) = \frac{n-N_u}{n}$$

Substituting this estimate into equation **3.2.1** affords the following tail estimator for F:

$$\hat{F}(z) = 1-\frac{N_u}{n}\left(1+\hat{\xi}\frac{z-u}{\hat{\beta}}\right)^{\frac{-1}{\hat{\xi}}}, \quad z > u.$$

Setting this expression equal to q gives an estimate for the q^{th} quantile

$$z_q = u+\frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{1-q}{N_u/n}\right)^{-\hat{\xi}}-1\right)$$

3.2.3 Implementation and Results

McNeil and Frey propose implementing their method in a two-stage approach which can be summarised as follows:

1) Fit a GARCH-type model to the returns data, and fit estimates $\hat{u}_{t+1}$ and $\hat{\sigma}_{t+1}$ using the fitted model.
2) If x_t represents the realised return on day t, calculate the model residuals as

$$\left(z_{t-n+1},\ldots,z_t\right)=\left(\frac{x_{t-n+1}-\hat{u}_{t-n+1}}{\hat{\sigma}_{t-n+1}},\ldots,\frac{x_t-\hat{u}_t}{\hat{\sigma}_t}\right),\ 0<n<t$$

Consider the residuals to be a realisation of the white noise process (Z_t), and use EVT to model the tail of $F_Z(z)$. Use this EVT model to estimate z_q for $q>0.95$.

McNeil and Frey tested their method using a similar backtesting technique to that used by Danielsson and De Vries. The size of the data set used to fit the GARCH and AR(1) models, and determine the realisations of the white noise process was 1000 days, and the method was tested on individual assets (in contrast to Danielsson and De Vries' test on portfolios of assets). The return series of the following assets were examined: S&P, DAX, BMW, USDGBP, and gold. VaR was calculated at the 95%, 99%, and 99.5% confidence levels for the purposes of the backtest. The VaR estimation techniques that were tested were the conditional and unconditional EVT methods, as well as methods in which the white noise process is assumed to have a particular distribution, i.e. normal (as in RiskMetrics or Student's t (fitting a heavy tailed distribution to the data). The results are summarised in Table 3.2.

Table 3.2: McNeil and Frey Backtesting Results

	S&P	DAX	BMW	USDGBP	Gold
Length of Test	7414	5146	5146	3274	3414
0.95 Quantile					
Expected	371	257	257	164	171
Conditional EVT	366	258	261	151	155
Conditional Normal	384	238	210	169	122
Conditional t	404	253	245	186	168
Unconditional EVT	402	266	251	156	131
0.99 Quantile					
Expected	74	51	51	33	34
Conditional EVT	73	55	48	35	25
Conditional Normal	104	74	86	56	43
Conditional t	78	61	52	40	29
Unconditional EVT	86	59	55	35	25
0.995 Quantile					
Expected	37	26	26	16	17
Conditional EVT	43	24	29	21	18
Conditional Normal	63	44	57	41	33
Conditional t	45	32	18	21	20
Unconditional EVT	50	36	31	21	11

In the analysis of their results, McNeil and Frey point out that in 11 out of 15 cases, their conditional EVT approach is closest to the mark. Application of a binomial test to determine the success of the quantile estimation methods, based on the number of violations, showed that on no occasion did their method fail. The conditional normal approach was found to fail 11 times, and unconditional EVT, 3 times. The unconditional EVT estimate cannot respond quickly to changing volatility and tends to be violated several times in a row in stress periods.[24d] Their overall conclusion can be summarised by the following quote: "approaches to tail estimation which ignore the conditional heteroskedasticity exhibited by most financial return series are not suitable for VaR calculation."[24d]

Chapter 4

Application of EVT-Based VaR Calculation Methods to the South African Market

This section of the report describes the investigation into the applicability of the EVT-based VaR calculation methods discussed thus far to the South African market. The published studies demonstrate that EVT-VaR performs well when tested on return series generated by US and European financial assets, but this observation is no guarantee of applicability to other markets, each of which have unique characteristics. With this point in mind, it was decided to conduct the study by testing the EVT-VaR calculation methods on return series generated by South African financial assets. The backtesting method used was based upon that employed by Danielsson and De Vries,[24b] and thus required a portfolio of stocks to form the basis of the analysis. The selection of these stocks is described in the following section. Initially, the method of historical simulation was tested, as it is a popular VaR estimation technique in South Africa, and the results of such a test would provide a benchmark against which to compare the EVT-based methods. Next, the method of Danielsson and De Vries was tested using the same selection of stocks, and finally, McNeil and Frey's approach was tested using the same method. The details concerning the implementation of these tests, as well as the results of the analyses are described in subsequent sections of this report.

4.1 Selection of Stocks

The backtesting method employed by Danielsson and De Vries uses a data set of 1500 consecutive trading days to estimate VaR, plus an additional 1000 consecutive trading days data to allow VaR to be calculated 1000 times (the key step in the backtest procedure). Therefore, a total of 2500 trading days of data was required for each stock that was going to form the basis of the portfolio analysis. This number corresponds to approximately 10 years of data. Thus the first step in the selection of the stocks for the study was to determine which companies had listed on the JSE prior to, or during 1991. A total of 132 such stocks was found. With a reasonable number of stocks with the required amount of data available, the final stage in the selction process was to determine which of these would be elements of a portfolio that could be considered as "representative". It was decided to examine the December 2000 compositions of unit trusts in the South African 'General Equity' category to provide guidance in this regard. One of the initial analyses involved determining, for each unit trust, what proportion of its stocks was from the pre-1991 list of stocks. The results of this analysis are presented in Table 4.1 (only the first 20 unit trusts are shown).

Table 4.1: Proportion of Unit Trusts' Compositions from pre-1991

Fund Name	Number of Stocks in Fund	Number of Pre 1991 Stocks in Fund	%
RMB High Tide	16	13	81.25
Investec Commodity R	25	19	76.00
Sage Resources	20	15	75.00
Metropolitan Resources	22	16	72.73
NIB Mining & Resource	22	16	72.73
Fedsure Wealth Specialist	29	21	72.41
FNB Growth	35	25	71.43
Sanlam Value	35	25	71.43
PSG Selective Top 40	24	17	70.83
Afr'Harvest Aggressive Value	20	14	70.00
Marriott Dividend Growth	20	14	70.00
Old Mutual Mining	16	11	68.75
Allan Gray Equity	59	40	67.80
Commercial Bank of Namibia	27	18	66.67
Old Mutual Gold	12	8	66.67
RSA Equity R	79	52	65.82
Liberty Value A	35	23	65.71
RMB Balanced	35	23	65.71
Metropolitan Namibia Pru Mgd	32	21	65.63

The subsequent analysis involved determining, for each pre-1991 stock, how often it occurred in the 10 unit trusts with the highest proportion of pre-1991 stocks. Table 4.2 shows the results for the first 20 stocks.

Table 4.2: Frequency of occurrence in top ten unit trusts

Code	Short Name	Number of Top 10 Unit Trusts in Which Share Occurs
DBR	DEBEERS	9
SOL	SASOL	9
AMS	AMPLATS	7
ASA	ABSA	6
GMF	GENCOR	6
SAP	SAPPI	6
IMP	IMPLATS	5
NED	NEDCOR	5
NHM	NORTHAM	5
SAB	SA-BREWS	5
SBC	SBIC	5
VNF	VENFIN	5
AIN	AVMIN	4
BAW	BARWORLD	4
BOE	BOE	4
GFI	GFIELDS	4
HAR	HARMONY	4
HVL	HIVELD	4

Another analysis focussed on the top 3 3-year and top 3 1-year performing unit trusts. In particular, the pre-1991 stocks that were common to the top 3 3-year performing unit trusts were identified. A similar analysis was carried out on the top 3 1-year performing unit trusts. The results of these analyses are displayed in Tables 4.3 and 4.4.

Table 4.3: Pre-1991 Stocks Common to Top 3 1-Year Performing Unit trusts

Code	Short Name
DBR	DEBEERS
SOL	SASOL
SBC	SBIC
AVI	A-V-I

Table 4.4: Pre-1991 Stocks Common to Top 3 3-Year Performing Unit trusts

Code	Short Name
SOL	SASOL
AMS	AMPLATS
NED	NEDCOR
SBC	SBIC
BOE	BOE
RCH	RICHEMONT
DDT	DIDATA
INT	INVSTEC
MET	METLIFE
FDS	FEDSURE
RLO	REUNERT

The main aim of the stock selection process was to obtain a portfolio of stocks that could be considered as a representative portfolio, as well as having the required amount of data available. The decision concerning what constitutes a representative portfolio is highly subjective, but it is hoped that a selection based on careful analysis will be considered as being reasonable. The stocks that were selected were chosen on the basis of the above analysis, as well as personal judgement. The selection is displayed in Table 4.5.

Table 4.5: Stocks Selected to Form the Basis of Portfolio Analysis

Code	Short Name
AMS	AMPLATS
BOE	BOE
DDT	DIDATA
INT	INVSTEC
LGL	LIBERTY
NED	NEDCOR
RCH	RICHEMONT
SBC	SBIC
SOL	SASOL

The following sections of the report describe the backtests of the VaR estimation methods discussed earlier. All analyses in the study were performed by algorithms written in the MATLAB programming environment.[26] The code for all the algorithms are included as appendices. MATLAB was chosen in preference to a language such as Visual Basic for Applications (VBA) because of its superior computational speed, and convenient handling of matrices.[27] As an example of how VBA and MATLAB compare, the historical simulation algorithm was coded in both VBA and MATLAB. After 24 hours, the VBA procedure had completed the analysis for only 200 randomly weighted portfolios, whereas MATLAB only required approximately 35 minutes to complete the analysis for 500 random portfolios.

4.2 Historical Simulation

The backtest of the historical simulation method of VaR calculation was carried out exactly as is described by Danielsson and De Vries (see Chapters 2&3).[24b] Essentially, the daily returns for the 9 stocks comprising the hypothetical portfolio were arranged into a 2486 x 9 matrix. A random 9 x 500 matrix, whose columns summed to one, was generated, which represented 500 random portfolio weights. Multiplication of these two matrices afforded a 2486 x 500 matrix, which represented 2486 days of returns for 500 different portfolios. This matrix formed the core of the analysis. Beginning at day 1500, for each portfolio, the previous 1500 days returns were used to compute VaR estimates (as discussed in Chapter 2) for various confidence levels. These VaR estimates were compared to the realised return on the following day. This was repeated 986 times. The results of the analysis are summarised in Table 4.6.

Table 4.6: Average Number of Violations of Historical Simulation Estimates

Confidence Level (%)	Expected	Observed	Standard Deviation
95	49.35	119.28	5.95
97.5	24.68	70.52	5.79
99	9.87	30.89	2.47
99.5	4.94	20.07	1.67
99.75	2.47	9.48	1.02
99.9	0.99	4.73	0.93
99.95	0.49	2.96	0.21

Based on the data used, the results clearly show how inadequate the method of historical simulation is, with the VaR estimates being violated approximately 3 times more frequently as expected. When tested in developed markets (e.g. US, Europe), historical simulation has been shown to offer reasonable VaR estimates for at least the 95% confidence level. That is clearly not the case for the SA market, and demonstrates the fact that the South African market has its own characteristics and peculiarities which need to be taken into account. It also shows that companies allocating risk capital based on estimates such as these are taking inadequate steps to protect themselves against extreme market risk, i.e. risk due to extreme market movements.

4.3 Danielsson and De Vries' Tail Estimator

The backtest of Danielsson and De Vries method was carried out as in the historical simulation method, except that VaR at each step was calculated using the tail estimator proposed by Danielsson and De Vries. The results of the backtest are displayed in Table 4.7.

Table 4.7: Average Number of Violations of Danielsson and De Vries Estimates

Confidence Level (%)	Expected	Observed	(Std Deviation)
95	49.35	127.31	8.18
97.5	24.68	72.98	6.55
99	9.87	29.13	2.35
99.5	4.94	14.85	1.94
99.75	2.47	6.87	1.43
99.9	0.99	2.68	0.59
99.95	0.49	1.30	0.53
99.9725	0.25	0.92	0.27
99.99	0.10	0.08	0.27
99.9955	0.05	0.00	0.00

It is clear that up to the 99% confidence level, the number of violations of the Danielsson and De Vries unconditional VaR estimates, and that of historical simulation (Table 4.6) are essentially the same. This similarity is not surprising, since the method of Danielsson and De Vries involves fitting a smooth curve through the tail of the empirical distribution function, and thus in a region where there is a higher concentration of empirical observations, the two methods are likely to have results that coincide. One would expect the two methods to diverge where data is sparse, the domain of interest in EVT. This phenomenon is exactly what is observed at confidence levels higher than 99%.

In order to assess fully the utility of this method, some additional analysis of the data was undertaken. Since the backtest involved calculating VaR approximately 1000 times for 500 different portfolios (i.e. 500 000 VaR calculations), it was clearly impossible to analyse the data in every single VaR calculation. It was decided to select a single vector of portfolio weights, and to examine the return series of this portfolio in two different 1500 trading day windows, labelled **a** and **b** in the figures below. The windows of portfolio returns were obtained by generating 2500 days of returns via the multiplication of the vector of portfolio weights with the matrix of stock returns, and selecting two starting points within this set of data from which to select 1500 days of data. In particular, window **a** was chosen to begin at return number 70 in the total of 2500, while window **b** was chosen to begin at return number 500. The selection of the two windows was arbitrary, although they do exhibit contrasting behaviour in terms of volatility. The return series from window **a** shows little change in volatility, whereas the series from window **b** displays large jumps, or spikes, corresponding to a period of high volatility. The two return series are shown in Figures 4.1 and 4.2 respectively.

Figure 4.1: Return Series Derived from Window **a**

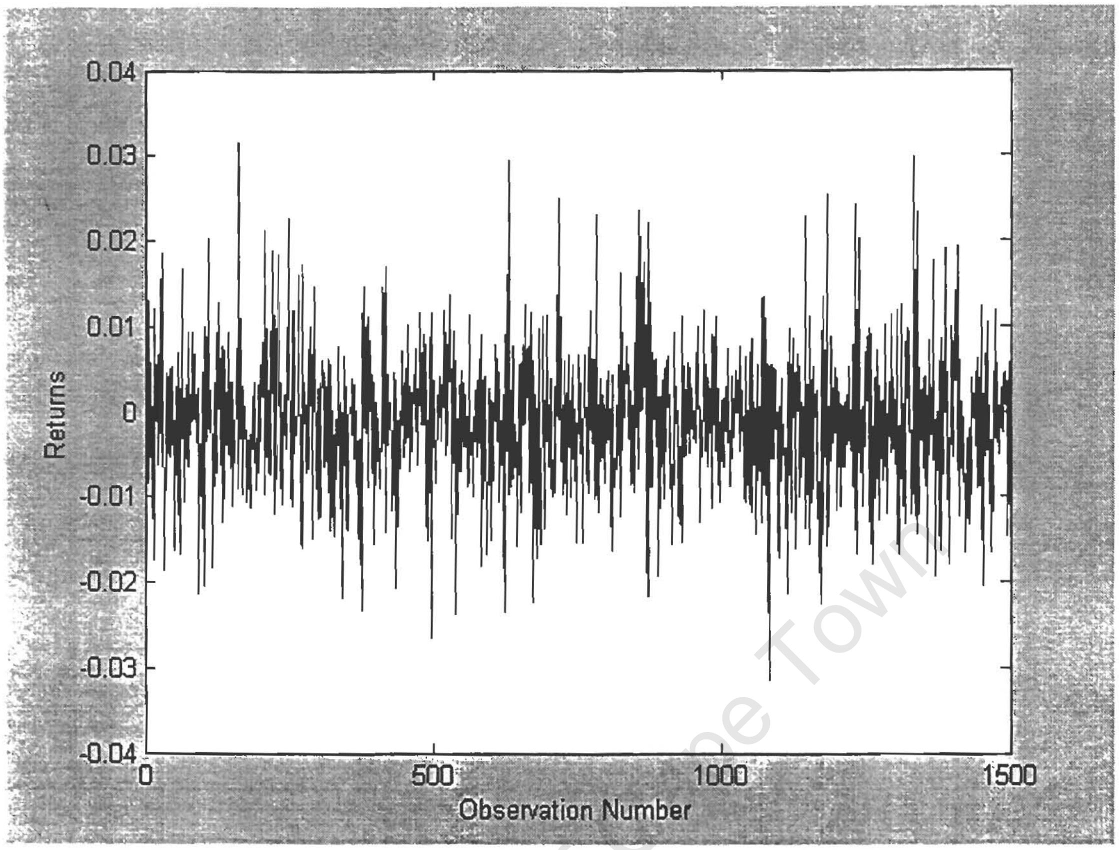
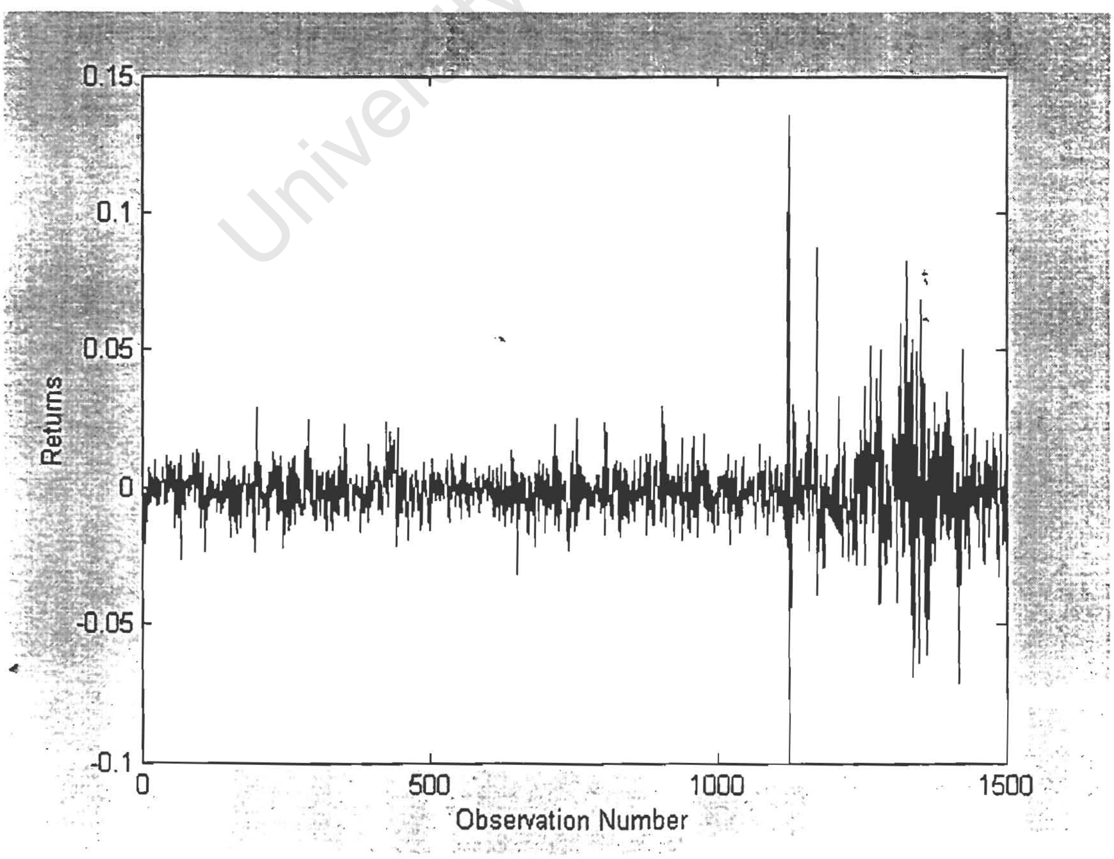


Figure 4.2: Return Series Derived from Window **b**



The purpose of this additional data analysis was simply to give an indication of the mechanics of the method, and not to provide any statistical inference.

One of the assumptions made in the derivation of the Danielsson and De Vries tail estimator is that the data are drawn from a heavy tailed distribution. A simple graphical technique which can be used to test the validity of this assumption is known as a *quantile-quantile plot* (*QQ-plot*). If a sample of size n is ordered as

$$X_{n,n} \leq \ldots \leq X_{1,n}$$

then the set of points that comprise the QQ-plot is defined as

$$\left\{ \left(X_{k,n}, F^{-1}\left(\frac{n-k+1}{n+1} \right) \right) : k = 1 \ldots n \right\}$$

This technique simply requires that the quantiles (or inverses) of the empirical distribution function and the assumed distribution function are plotted against each other. Thus if the distributional assumption is valid, the plot should be approximately linear. For the purposes of detecting heavy tailedness in financial return series, F is assumed to be a normal distribution function. If the plot curves down at the left and/ or up at the right, then it may be assumed that the return series has heavier tails than normal.[28]

QQ-plots were generated for the return series of windows **a** and **b**, and are displayed in Figures 4.3 and 4.4. A normal distribution with mean and standard deviation estimated from the data was used. In both cases, the plots clearly display the heavy-tailed nature of the return series, this feature being even more pronounced in the case of window **b**.

Figure 4.3: QQ-Plot (Window **a**)

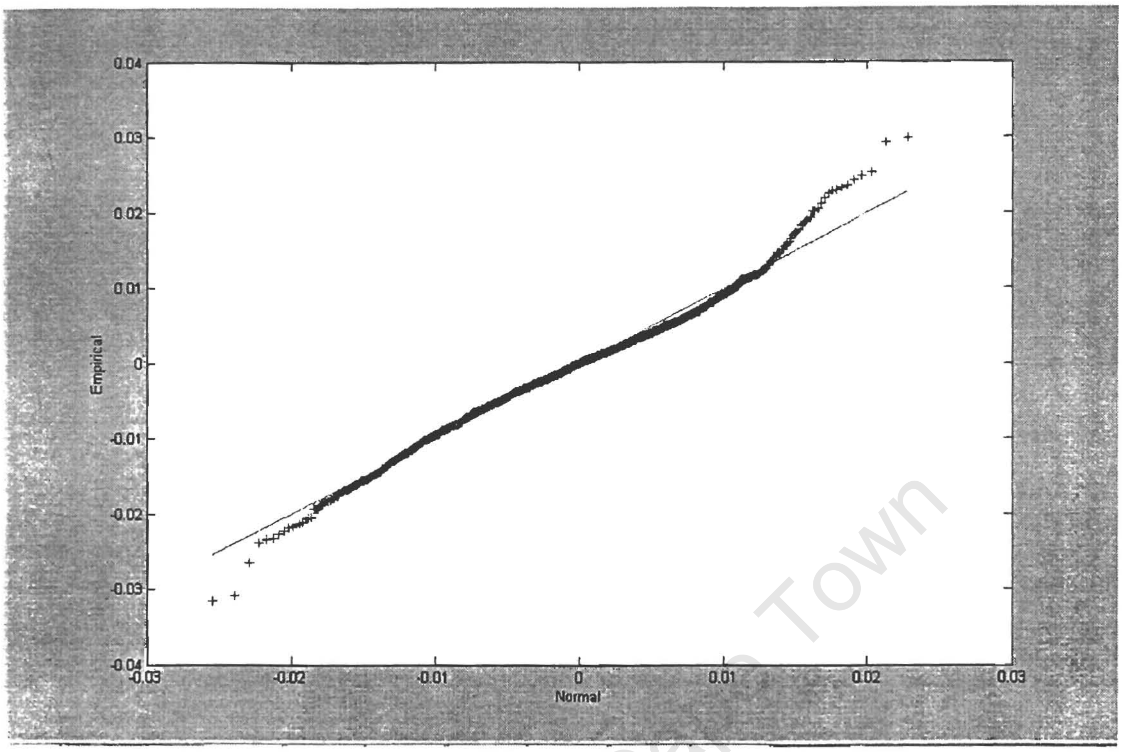
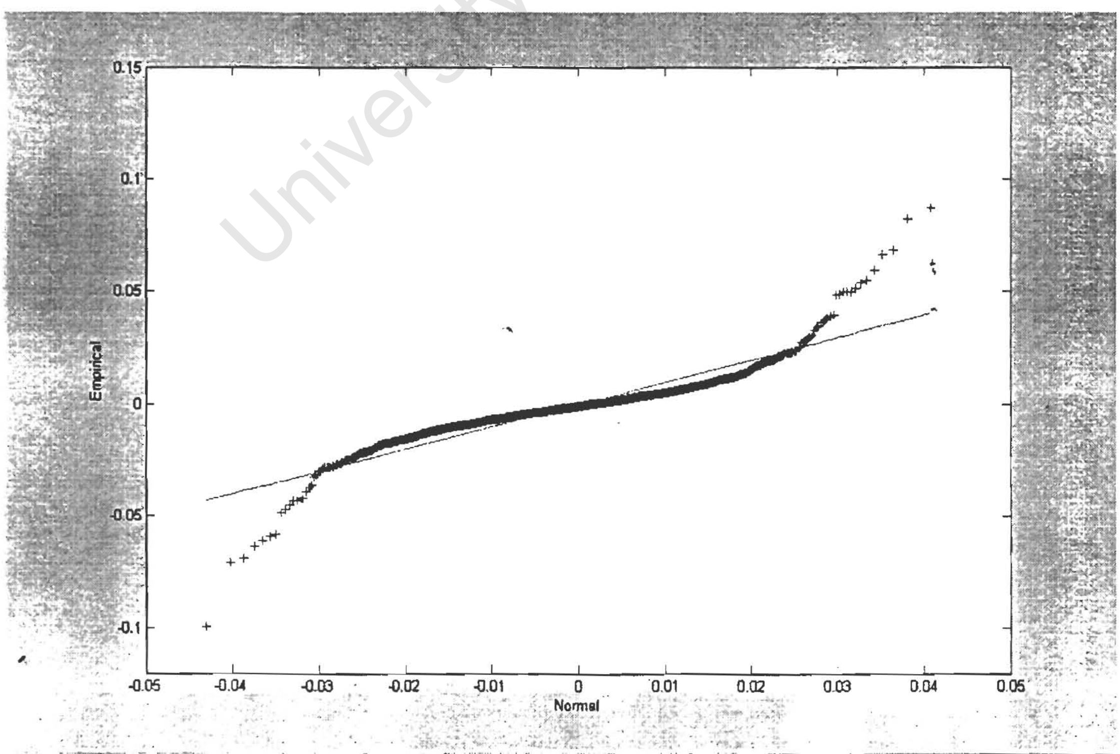


Figure 4.4: QQ-Plot (Window **b**)



Since the essence of Danielsson and De Vries' method is the fitting of a smooth curve through the tail of the empirical distribution function, it was decided to plot the fitted curves for the windows **a** and **b** to obtain a sense of how well Danielsson and De Vries' tail estimator actually approximates the tails. Figures 4.5 and 4.6 show the tail of the empirical distribution function for the two windows, along with fitted tails. In the case of window **b**, the fit appears to be satisfactory, but less so in the case of window **a**. In order to obtain a quantitative indication of how well the fitted curve approximates the tail, the *Kolmogorov-Smirnov (K-S) goodness-of-fit test* was applied to both plots. The K-S test involves comparing the largest absolute standard deviation of the empirical distribution function from the fitted curve with tabulated values.[27] Formally,

$$D(n) = \max_x |F_n(x) - F(x)|$$

where $F_n(x)$ denotes the empirical distribution function, and $F(x)$ the fitted curve. As indicated by the notation, the value of the statistic is dependent on the sample size. At the 99% confidence level, and for $n > 50$, the tabulated value of D is[30]

$$\frac{1.63}{\sqrt{n}}.$$

Since the K-S test is generally used to test the goodness-of-fit of an entire distribution function, the tail region on the probability axis for each plot was scaled to the interval $[0,1]$, via the transformation

$$T(x) = \frac{x-a}{b-a}$$

where a is the lowest probability in the area of interest, and b the highest (i.e. 1). This technique was used in the absence of any knowledge of a goodness-of-fit test dedicated to the tails.

The sample size when considering the tails was 150, which translated to a tabulated D value of 0.1331. The computed D value for window **a** was 0.1018, and that of **b** was 0.0657. Both values are less than the tabulated D, thus indicating an acceptable goodness-of-fit in both cases, although the D-values and the graphical plots indicate that the fit is better in the case of window **b** than in **a**.

Figure 4.5: Empirical Distribution Function With Fitted Tail (Window **a**)

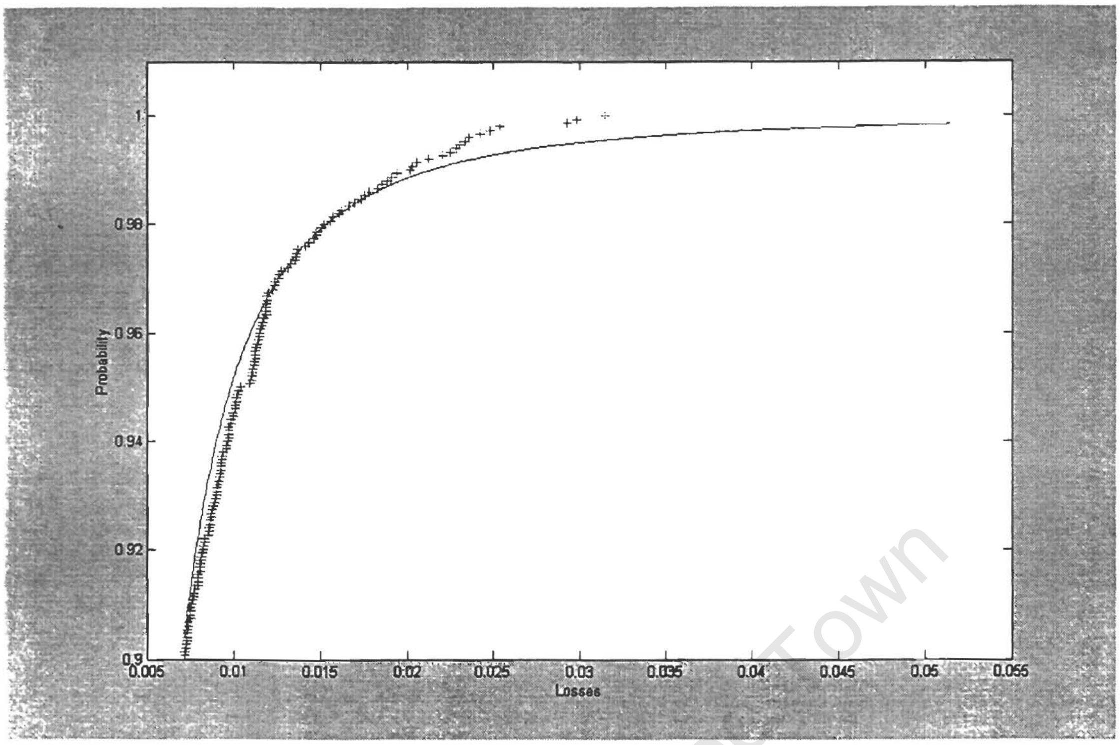
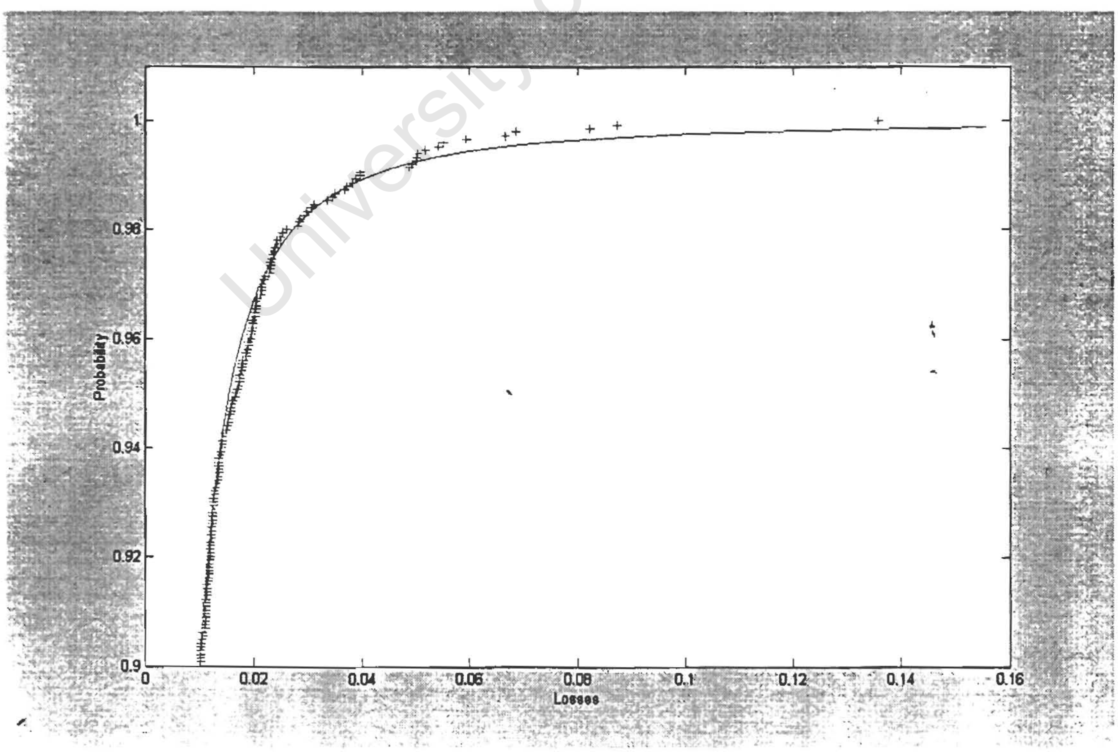


Figure 4.6: Empirical Distribution Function With Fitted Tail (Window **b**)



The final aspect of Danielsson and De Vries' method that was examined was the actual VaR estimates that were produced. Table 4.8 shows the VaR estimates that were calculated using the data from windows **a** and **b**. Also included are VaR estimates calculated using the same data, but via the method of historical simulation.

Table 4.8: VaR Estimates Produced By Danielsson and De Vries' (D&D) Tail Estimator

	Window a		Window b	
	HS	**D&D**	**HS**	**D&D**
95	0.0109	0.0098	0.0170	0.0155
97.5	0.0137	0.0137	0.0234	0.0238
99	0.0203	0.0214	0.0397	0.0418
99.5	0.0234	0.0299	0.0541	0.0640
99.75	0.0254	0.0418	0.0687	0.0981
99.9	0.0298	0.0651	0.0873	0.1724
99.95	0.0315	0.0910	0.1358	0.2642
99.9725	-	0.1272	-	0.4047
99.99	-	0.1981	-	0.7113
99.9955	-	0.277	-	1.0897

Up until the 99% confidence level, historical simulation and Danielsson and De Vries' method afford similar VaR estimates, and then diverge at higher confidence levels, as discussed earlier. The most noteworthy (and important) feature of these numbers is the very high VaR estimates produced by Danielsson and De Vries' method, particularly in the case of window **b**. This, coupled with the fact that not once was the highest confidence level VaR estimate ever violated, suggests that Danielsson and De Vries' method may be prone to *overestimating* VaR in the extreme tail region, for this data set. On the whole, the VaR estimation procedure does not produce consistent results – VaR is underestimated at the lowest confidence levels, while overestimated at the highest. Thus not too much emphasis can be placed on any reasonable performance in between these two extremes (e.g. at the 99.99% level).

4.4 McNeil and Frey's Tail Estimator

The method used to backtest McNeil and Frey's VaR estimation procedure was the same as that used in the historical simulation case, and in the test of the Danielsson and De Vries estimator to ensure consistency throughout the report. This backtest approach differs to that adopted by McNeil and Frey in that McNeil and Frey use a data window of 1000 trading days to estimate VaR and in addition, only consider return series generated by single assets, as opposed to returns on a portfolio. The latter difference is worthy of further discussion. McNeil and Frey state that "the method obviously also applies to the time series of profits and losses generated by portfolios of financial instruments and can therefore be used for the estimation of market risk measures in a portfolio context".[24d] In a separate paper,[9] McNeil gives a brief overview of *multivariate* extreme value theory (MEVT), and outlines a simple bivariate POT model. However, it is pointed out that parametric models of this kind are viable only in a small number of dimensions. "In very high dimensions, there are simply too many parameters to estimate...In such situations, collapsing the problem to a univariate problem by considering a whole portfolio of assets as a single risk and collecting data on a portfolio level seems more realistic".[9] This approach of 'collapsing' the portfolio has been the approach adopted throughout the study.

The MATLAB code used for the backtest is included in Appendix C. This backtest was computationally the most complex of the three backtests to execute, as each VaR calculation involved the estimation of GARCH(1,1) parameters, as well as determination of parameters for the fitted GPD. In terms of computation time, the estimation of the GARCH parameters via the maximum likelihood method was the most demanding. Since this aspect of McNeil and Frey's method plays a crucial role, it warrants further discussion.

In very basic terms, maximum likelihood estimation involves selecting a particular distribution for the data from which the parameters are to be estimated. One derives a function of the parameters that gives a measure of the probability (or likelihood) of the data occurring. The maximum likelihood estimate of the parameters consists of those parameter values that maximise this function. The data used to estimate GARCH parameters are obtained by fitting an econometric model to the return data, and calculating a realisation of the heteroskedastic noise process. In McNeil and Frey's model, this would involve estimating ϕ, and obtaining the noise process as $X_t - \hat{\phi}X_{t-1} = \hat{\varepsilon}_t$. The set $\{\hat{\varepsilon}_t\}$ is used to estimate the GARCH parameters. McNeil and Frey refer to their maximum likelihood procedure as the *pseudo*-maximum likelihood method (PML). This fact is noted because the data used to estimate the GARCH parameters are assumed to be normally distributed for the purpose of parameter estimation, "a distributional assumption we don't necessarily believe".[24d] However, McNeil and Frey state that the PML method delivers reasonable parameter estimates. Thus if the data are assumed to be normally distributed, and subject to a volatility updating scheme

$$\sigma_i^2 = \kappa + \alpha\sigma_{i-1}^2 + \beta\varepsilon_{i-1}^2$$

then the probability density for the i^{th} observation ε_i is

$$\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(\frac{-\varepsilon_i^2}{2\sigma_i^2}\right)$$

The probability density of the n observations occurring in the order that they are observed is

$$\prod_{i=1}^{n}\left[\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(\frac{-\varepsilon_i^2}{2\sigma_i^2}\right)\right]$$

The best estimate of the parameters κ, α and β are the values that maximise this expression. This is equivalent to finding parameters that maximise the logarithm (ignoring constant multiplicative factors) i.e.

$$\sum_{i=1}^{n}\left[-\ln\left(\sigma_i^2\right) - \frac{\varepsilon_i^2}{\sigma_i^2}\right]$$

This expression is known as the *log-likelihood* function for the GARCH parameter estimation procedure. An important point to note is that the calculation of the conditional variances $\left\{\sigma_t^2\right\}_{t=1}^{T}$ from the GARCH equation requires pre-sample values for σ_0^2 and ε_0^2, i.e. values are needed to "jump start" the process. Bollerslev suggested setting both of these to the unconditional variance of the observed sample,[20] which is the approach adopted by the MATLAB GARCH parameter estimation algorithm.

The initial part of the MATLAB code used in this backtest is identical to that used in the previous method, *viz.* the setting up of a matrix of returns for 500 different portfolios. The other major components of the code are the estimation of the GARCH parameters and the fitting of the GPD. The GARCH parameter estimation was carried out by making use of a convenient MATLAB facility called the GARCH Toolbox. This allows one to specify an econometric model for the returns data, pass the returns data as an argument to the Toolbox, and obtain the estimated GARCH parameters as output. Attempts to implement the backtest procedure with the AR(1) process proposed by McNeil and Frey as the specified model resulted in frequent "stalling" during the maximisation of the likelihood function routine. According to the MATLAB GARCH Toolbox manual,[18] this stalling is indicative of an inappropriate model for the data. The author(s) suggest using the simpler default model without an AR(1) parameter

$$X_t = C + \sigma_t Z_t$$

(C constant) as it is adequate for most purposes. Empirical tests using the stock returns data at hand did indeed show that the GARCH estimation procedure using the simpler model specification was less prone to "stalling" than in the more complex AR(1) case. Since the backtest involved approximately 500 000 GARCH parameter estimations, it was decided to run the test with the simpler default model. We believe that this model still captures the essence of McNeil and Frey's method.

With the GARCH model parameters in hand, a realisation of the strict white noise process $\{Z_t\}$ could be calculated, as described in Chapter 3. The set of realisations was used in the next major component of the code, *viz.* the estimation of the parameters ξ and β (not to be confused with the GARCH parameter β) for the fitted GPD. As shown by Embrechts *et al.*,[31] the maximum likelihood estimates of ξ and β are given by

$$\hat{\xi} = \hat{\xi}(\tau) = \frac{1}{N}\sum_{i=1}^{N}\ln(1-\tau Y_i)$$

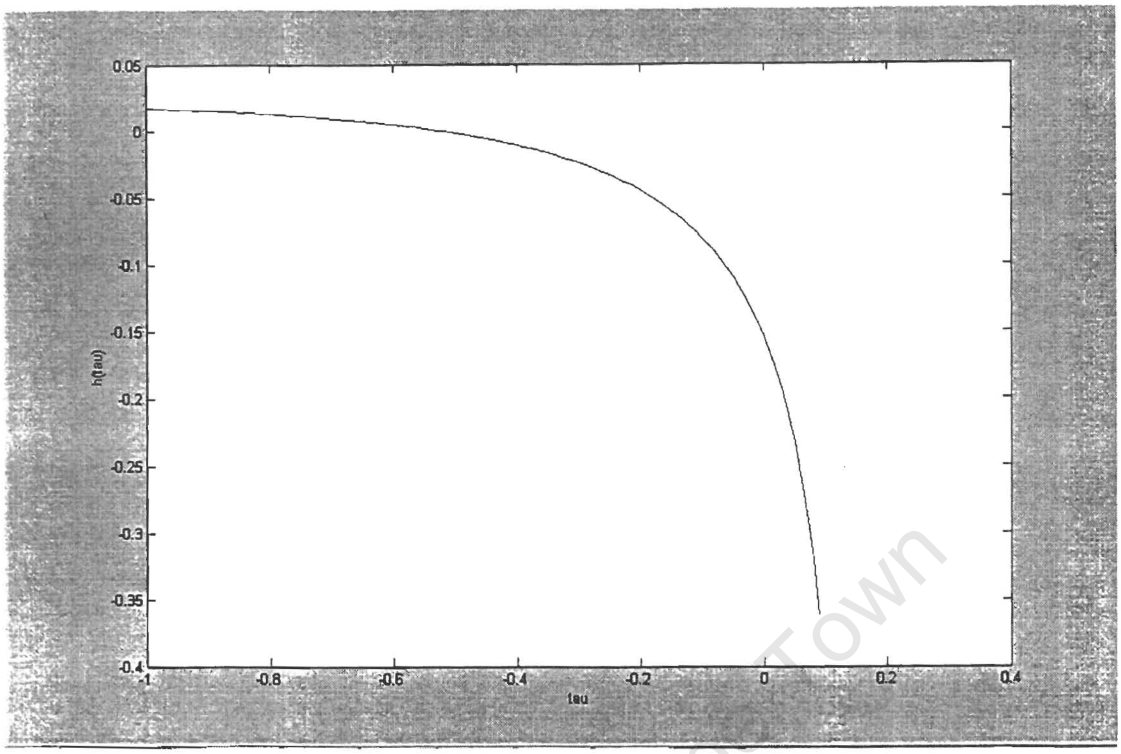
$$\hat{\beta} = \frac{-\hat{\xi}}{\tau}$$

where $Y_1,...,Y_N$ are the N excesses over a high threshold, and τ satisfies

$$h(\tau) = \frac{1}{\tau} + \frac{1}{N}\left(\frac{1}{\hat{\xi}(\tau)}+1\right)\sum_{i=1}^{N}\frac{Y_i}{1-\tau Y_i} = 0$$

Thus the key step in the estimation of ξ and β is the determination of τ_0 such that $h(\tau_0) = 0$. This root was found via Newton's method using MATLAB code modified from Borse.[32] In order to select an appropriate initial estimate for the Newton's method algorithm, a graph of $h(\tau)$ vs. τ was plotted using arbitrarily chosen data (Figure 4.7). Based on this plot, an initial estimate of τ = -0.05 was used in all of the backtest calculations. Out of a total of approximately 500 000 VaR estimation calculations, the Newton's method algorithm failed to converge to any unique root at most only 10 times.

Figure 4.7: Plot to Determine Initial Estimate for Newton's Method



With the GARCH and GPD parameters in hand, VaR values for various confidence levels were easily calculated based on the formulae described in Chapter 3. It is important to note, however, that in using the altered model specification for the returns data, the relationship between the quantile of the conditional distribution function of the returns, and that of the strict white noise process becomes

$$x_q = C + \sigma_{t+1} z_q$$

Also calculated were VaR estimates based on the assumption of the white noise process having a standard normal distribution. These estimates involved simply setting z_q equal to the corresponding quantile of the standard normal distribution function. The reason for this second set of calculations was to allow a backtest of a RiskMetrics™ type approach to VaR estimation. The results of the backtest are displayed in Table 4.9.

Table 4.9: Average Number of Violations of McNeil and Frey's VaR Estimates

		EVT-GARCH		Normal GARCH	
	Expected	**Observed**	**(Std Deviation)**	**Observed**	**(Std Deviation)**
95	49.35	77.91	4.43	69.77	5.07
97.5	24.68	42.708	3.24	47.02	3.24
99	9.87	21.55	2.20	31.63	2.64
99.5	4.94	12.156	1.95	24.91	2.29
99.75	2.47	6.394	1.35	20.16	2.19
99.9	0.99	2.778	0.99	15.55	2.36
99.95	0.49	1.79	0.74	12.66	2.31
99.9725	0.25	1.172	0.67	10.32	2.29
99.99	0.10	0.648	0.60	7.97	2.12
99.9955	0.05	0.446	0.55	6.58	1.93

The results clearly show the benefit of incorporating a volatility updating scheme into the calculation of VaR. At the 95% confidence level, the GARCH-based methods outperform the other two, with the RiskMetrics™-type approach providing the best results at this level. However, at higher confidence levels, the conditional normality assumption leads to serious under-predictions of VaR. As in the case of the analysis of Danielsson and De Vries' method, further data analysis was undertaken to obtain insight into the mechanics of McNeil and Frey's method, and to assist in the interpretation of the backtest results. The same two "windows" of data that were examined in the discussion of Danielsson and De Vries' method were used here, to allow a direct comparison.

Figures 4.8 and 4.9 show the noise process corresponding to the fitted model (innovations), the conditional volatility series, and the realisation of the white noise process (standardised innovations). The plots for window **b** are more interesting, as they clearly show how a period of large fluctuations in the return series is taken into account by the conditional volatility model, and how the white noise process is free of such spikes, the graphical plot indicating a constant variance process.

Figure 4.8: Conditional Standard Deviations and White Noise Process (Window **a**)

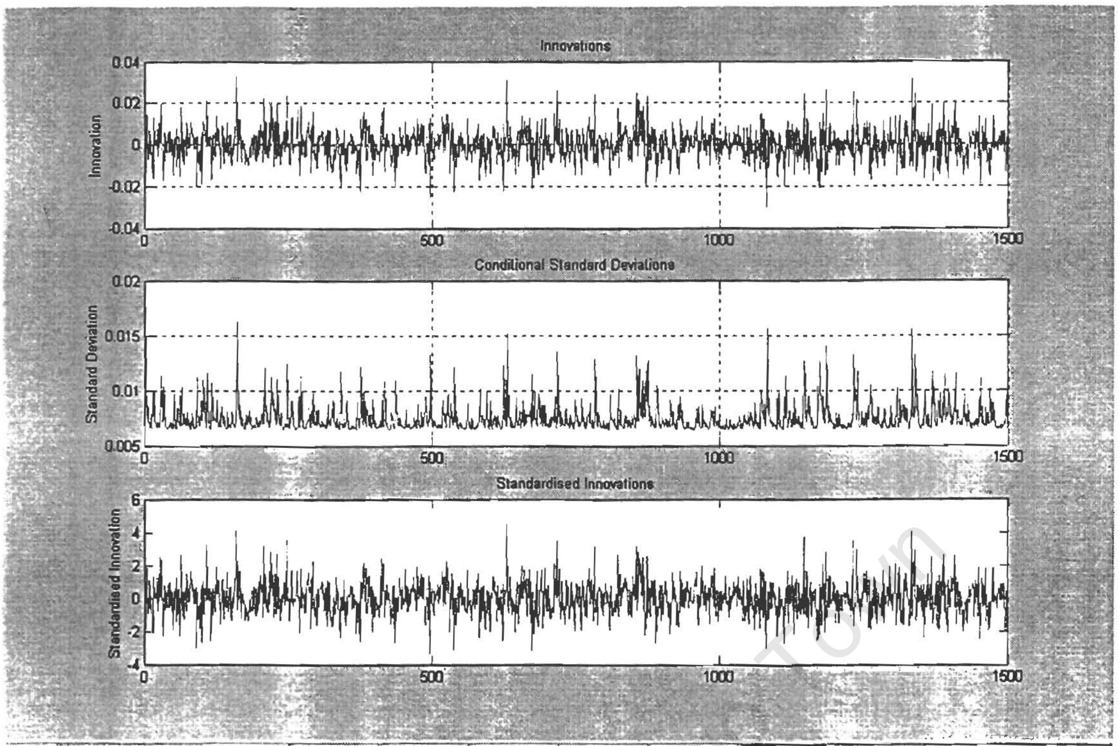
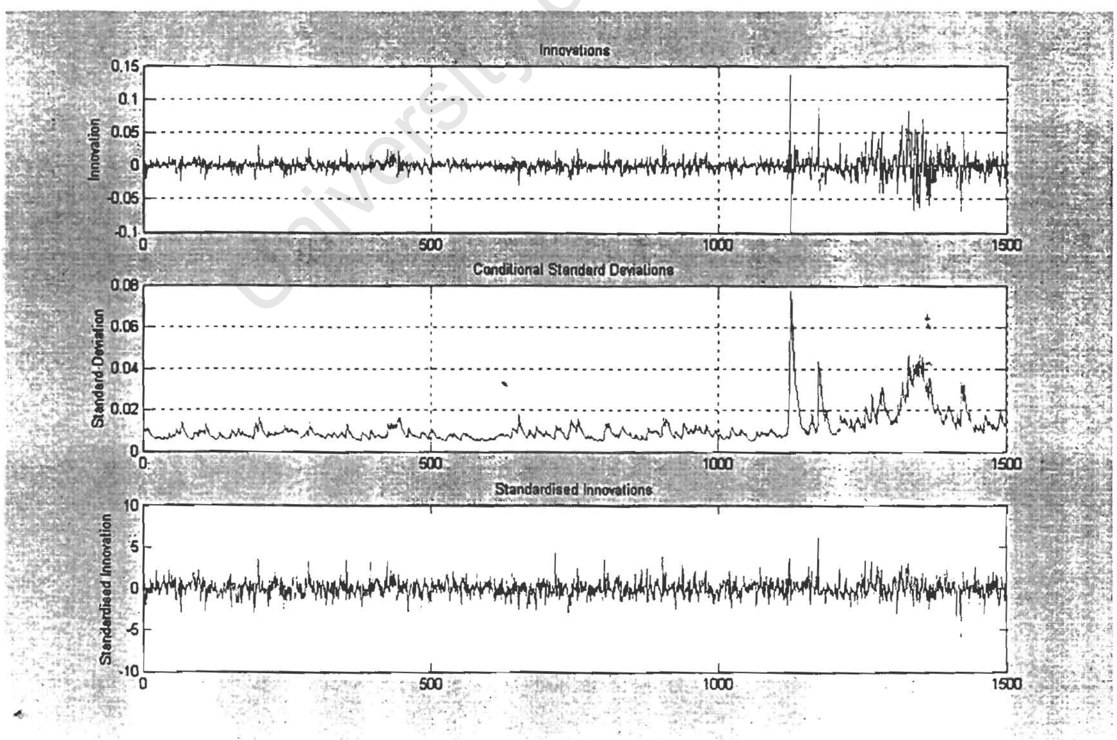


Figure 4.9: Conditional Standard Deviations and White Noise Process (Window **b**)



One of the assumptions made in the choice of the GPD was that the white noise process is heavy tailed. This property can clearly be seen to be the case for the two windows in the QQ-plots shown in Figures 4.10 and 4.11. Thus the assumption of conditional normality is unrealistic.

Figure 4.10: QQ-Plot (Window **a**)

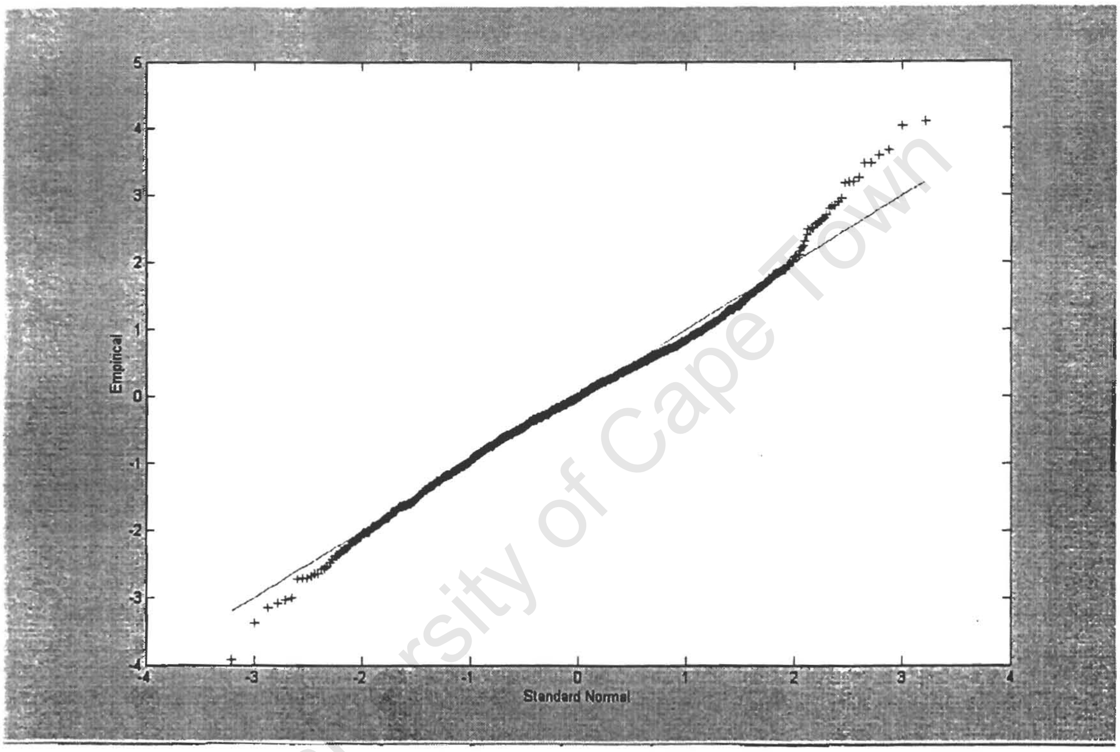
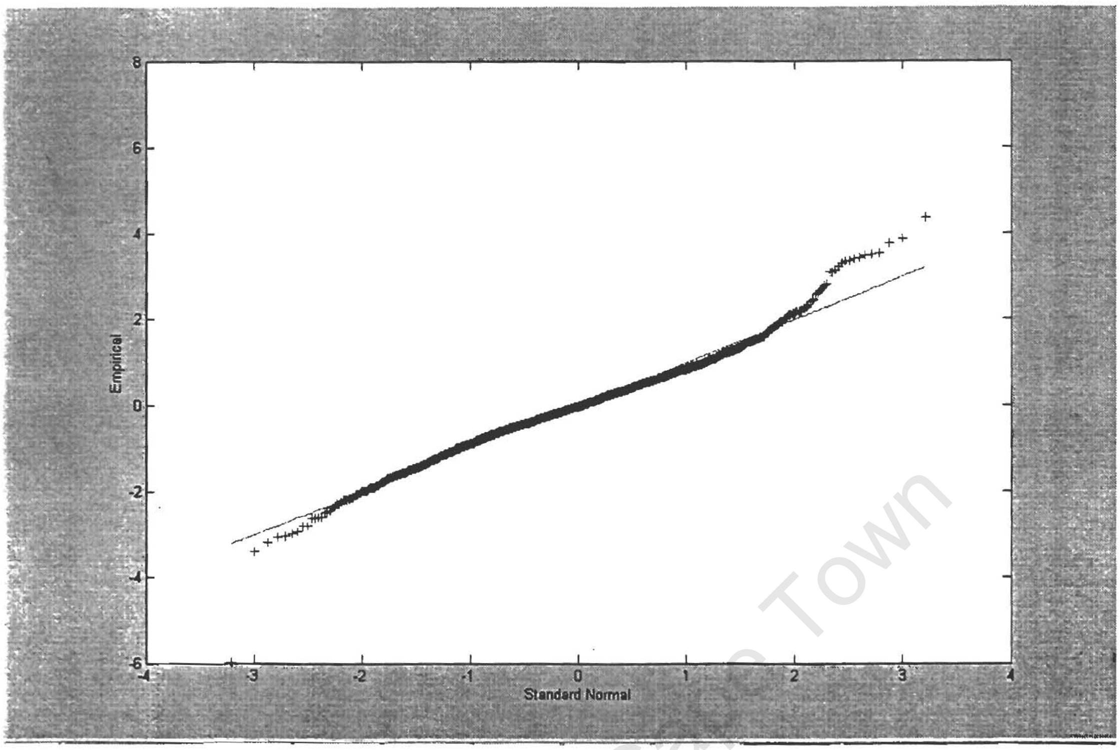


Figure 4.11: QQ-Plot (Window **b**)



As in the case of the VaR estimator of Danielsson and De Vries, an important step in McNeil and Frey's method involves the fitting of a smooth curve to a distribution function of extreme observations and thus it is important to examine how good the fit is to that subset. Figures 4.12 and 4.13 clearly show that the estimated GPD curve is a good fit to the empirical distribution function in both cases. In addition, application of the K-S goodness-of-fit test yields D values of 0.0455 and 0.0436 for windows **a** and **b** respectively, well below the tabulated value of 0.1331 (99% confidence level).

Figure 4.12: Empirical Excess Distribution Function and Fitted GPD (Window **a**)

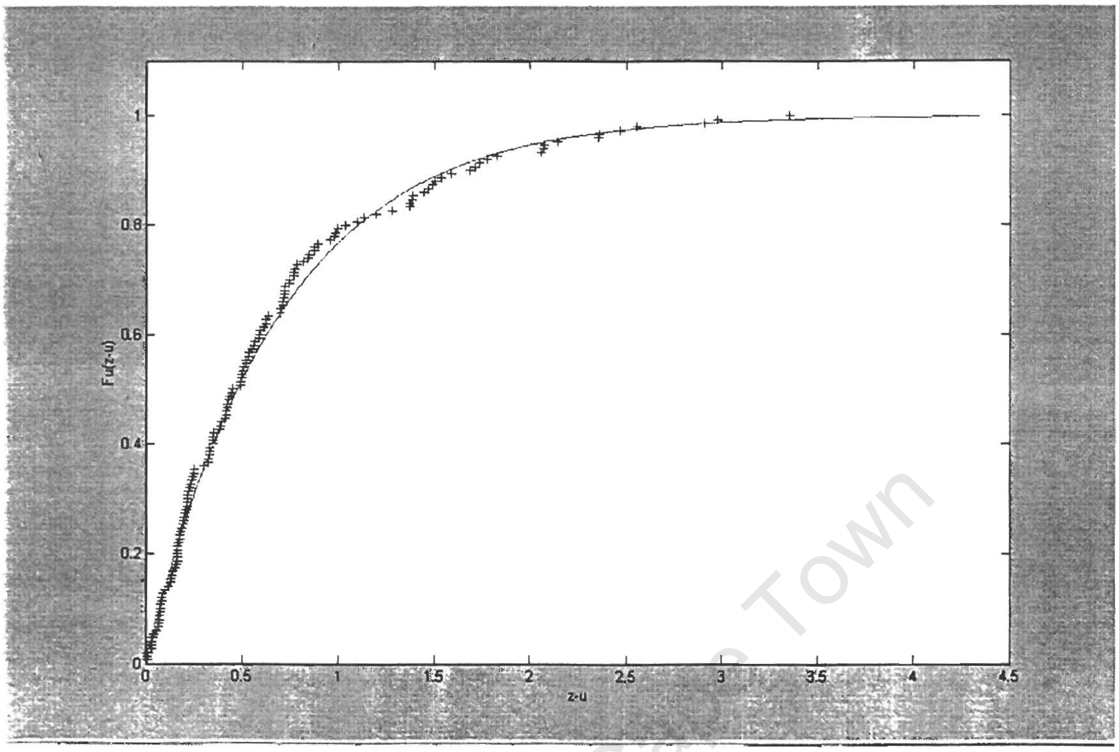
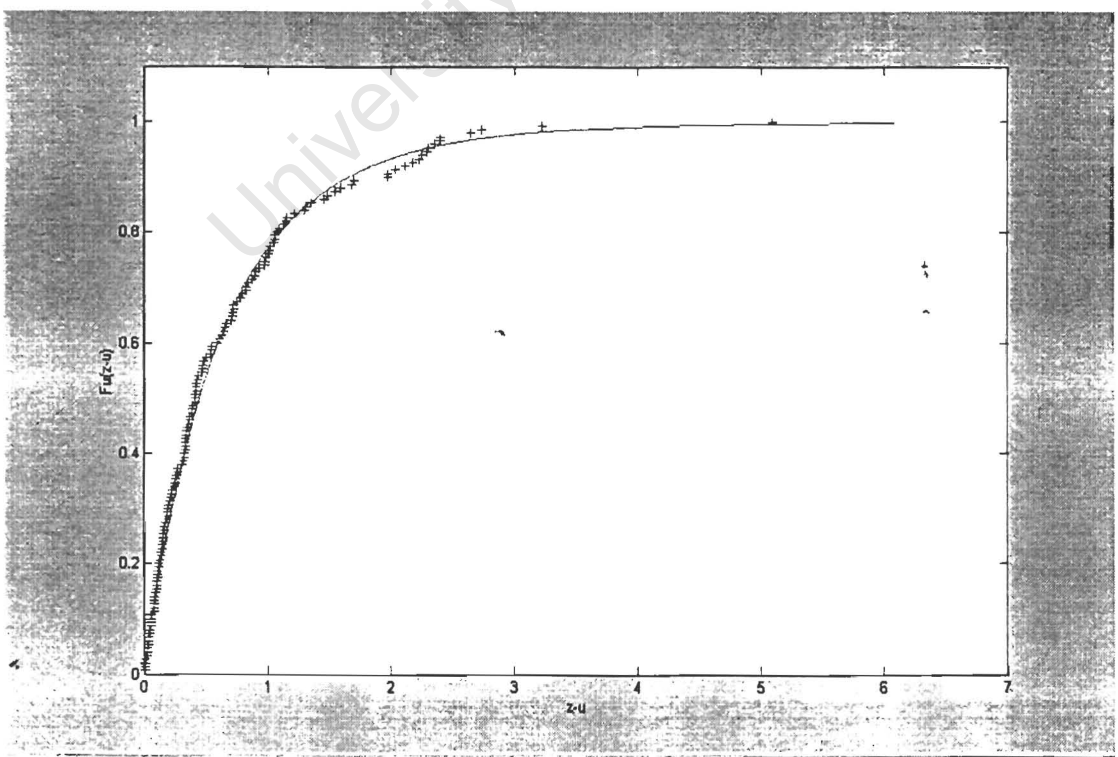


Figure 4.13: Empirical Excess Distribution Function and Fitted GPD (Window **b**)



The final aspect of McNeil and Frey's method that was examined was the actual VaR estimates themselves. Table 4.10 shows VaR estimates for various confidence levels using the data from windows **a** and **b**. Also included are the estimates provided by historical simulation and the method of Danielsson and De Vries , for comparative purposes. All three methods provide similar estimates at the lower confidence levels. At the highest levels, the large discrepancy between the values predicted by Danielsson and De Vries' estimator and that of McNeil and Frey is evident.

Table 4.10: VaR Estimates Produced By McNeil and Frey's, and Danielsson and De Vries' (D&D) Tail Estimator

	Window a			**Window b**		
	HS	**D&D**	**McNeil**	**HS**	**D&D**	**McNeil**
0.95	0.0109	0.0098	0.0095	0.0170	0.0155	0.0177
0.975	0.0137	0.0137	0.0126	0.0234	0.0238	0.0235
0.99	0.0203	0.0214	0.0168	0.0397	0.0418	0.0321
0.995	0.0234	0.0299	0.0200	0.0541	0.0640	0.0392
0.9975	0.0254	0.0418	0.0231	0.0687	0.0981	0.0470
0.999	0.0298	0.0651	0.0272	0.0873	0.1724	0.0584
0.9995	0.0315	0.0910	0.0304	0.1358	0.2642	0.0679
0.99975	-	0.1272	0.0335	-	0.4047	0.0783
0.9999	-	0.1981	0.0376	-	0.7113	0.0934
0.99995	-	0.2770	0.0407	-	1.0897	0.1060

In summary, based on the backtest results and the magnitudes of the VaR estimates provided by the methods, the method of McNeil and Frey performed the best in this study of the SA market. Even though McNeil and Frey's VaR estimate was violated on more occasions at the higher confidence levels than that of Danielsson and De Vries, we believe that this is simply a result of unrealistically large VaR estimates afforded by Danielsson and De Vries' method. On the whole, the method of McNeil and Frey appeared to be more consistent in its performance.

Even though the method of McNeil and Frey afforded the best results, the number of violations of the VaR estimates was much higher than the expected number of violations at all confidence levels. This is in contrast to the results of the backtests performed by both Danielsson and De Vries, and McNeil and Frey, who report very good agreement between the expected number of violations, and the realised number of violations. This unexpected phenomenon is indicative of a point raised earlier, that methods with proven success in foreign, developed markets are not necessarily appropriate for the South African market. The results obtained in this study suggest that historical simulation and the associated "tail smoothing" approach are not of great utility in the local market, whereas methods incorporating volatility updating show some promise. The results therefore also indicate that any attempts to obtain an improved VaR estimation method should have an approach similar to that of McNeil and Frey as a starting point. In this study, a very brief investigation was made into a possible modification of McNeil and Frey's method with the aim of obtaining a better agreement between the expected and realised number of violations of the VaR estimates. The two main features of McNeil and Frey's method are the fitting of a GPD to the tail of the white noise distribution function, and obtaining a one step ahead volatility forecast via a GARCH model. Any attempts to modify the method could therefore focus on one, or both of these aspects. The fact that the GPD curve appeared to be a very good fit to the data of windows **a** and **b** suggested that a possible source of improvement would be the use of an alternative GARCH model to provide the one step ahead volatility forecast. Some possible alternatives are discussed in the next section.

4.5 Asymmetric GARCH models

In the standard GARCH(r, m) model, negative and positive returns/ innovations of the same magnitude lead to the same conditional variance σ_t^2. However, empirical studies on financial return series have shown that they are characterised by increased conditional variance following negative shocks (bad news).[33] This observation has led to the development of what are termed "asymmetric GARCH models", which attempt to capture the asymmetric effect of news (or shocks) on the conditional variance. The common feature of models such as these is that a negative return/ innovation is given a higher weight in the calculation of the conditional variance. A study published by Engle and Ng in 1993 compares the effectiveness of several asymmetric GARCH models, and also introduces a diagnostic tool known as a news impact curve.[34] The news impact curve is simply a plot that shows how the predicted conditional variance varies as a function of returns/ innovations. This plot will be discussed further later.

After an examination of some of the asymmetric GARCH models available, it was decided to attempt a modification of McNeil and Frey's method, with the standard GARCH model replaced by an asymmetric GARCH model proposed by Glosten, Jagannathan, and Runkle in 1993 (GJR-GARCH).[35] This extension was chosen as it is a simple modification of the symmetric GARCH framework, and in addition, was found to be the best of the parametric models studied by Engle and Ng.[34] The GJR model corresponding to the symmetric GARCH(1,1) case is given by

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

$$\text{where } I_{t-1} = \begin{cases} 1 & \varepsilon_{t-1} < 0 \\ 0 & \varepsilon_{t-1} \geq 0 \end{cases}$$

In this model it is clear that the indicator function I_t serves to increase the predicted conditional variance following a negative shock, relative to that following a positive shock. An increase in the predicted conditional variance in McNeil and Frey's method may lead to a better agreement between the expected and realised violations of the VaR estimates, as this would lead to slightly higher VaR estimates. In order to implement a backtest with this method, it was necessary to make some modifications to the MALTAB code used to estimate the GARCH parameters. Firstly, an alternative, simpler MATLAB function called *ugarch* was used, as it was more amenable to modification than the functions used in the original backtest of McNeil and Frey's method. In addition, a time series model of

$$X_t = \sigma_t Z_t$$

was assumed (i.e. the RiskMetrics™ method) as it allowed the raw return series data to be used directly in the GARCH parameter estimation. The modified MATLAB code is provided in Appendix D. The modification entailed altering the expression for the conditional variance to include the extra term, and changing the call to the maximisation of the log-likelihood function in order to return the additional parameter γ.

Unfortunately, the use of the simpler GARCH parameter estimation function resulted in a much slower backtest algorithm that was extremely prone to stalling. As a result, only 3 portfolios (instead of 500) could be tested in the available time. Although no definitive conclusion can be made based on a sample as small as this, it is tempting to speculate that the use of GJR-GARCH would not result in an improved backtest result. An indication of this possible state of affairs is given by the fact that the numbers of violations of the modified procedure and the standard GARCH method are nearly identical for the same portfolios. To gain some insight into why this might be so, GJR and standard GARCH parameters were calculated using an arbitrarily chosen data window, and are presented in Table 4.11.

Table 4.11: Estimated Standard and GJR GARCH Parameters

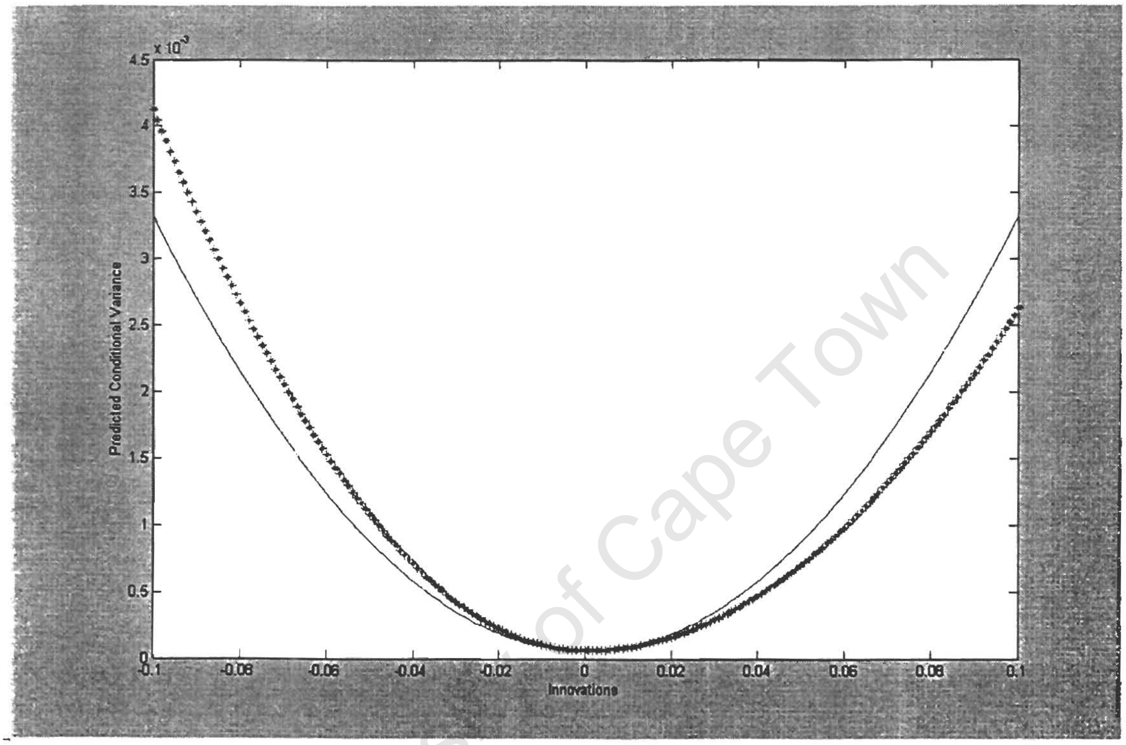
	Standard	GJR
κ	0	0
α	0.4215	0.4407
β	0.3269	0.257
γ	-	0.1497

An important point to note is that the value of β in the symmetric GARCH case is higher than that in GJR-GARCH, thus suggesting that GJR-GARCH would deliver a lower predicted conditional variance estimate following a positive shock than the standard GARCH model. This reductioncan be seen graphically via the news impact curve given in Figure 4.14. The curve was generated by setting σ^2 equal to the unconditional variance of the sample, and plotting σ_t^2 as a function of ε_{t-1}^2, using the estimated model parameters, i.e.

$$\sigma_t^2 = \alpha\sigma^2 + \beta\varepsilon_{t-1}^2 + \gamma\varepsilon_{t-1}^2 I_{t-1}$$

The corresponding symmetric GARCH plot is given by the solid curve. The figure clearly shows how the GJR model gives a lower predicted variance following a positive shock, and a higher following a negative shock, relative to the standard GARCH model. We propose that in order to improve the backtest results of McNeil and Frey's method, a GARCH model should be investigated that affords the same volatility estimate as the standard GARCH model following a positive shock, and a higher estimate following a negative shock.

Figure 4.14: News Impact Curves For Standard (Solid Curve) and GJR-GARCH (***) Models



Chapter 5

Conclusion

The results in this study indicate that historical simulation is an inadequate method for the calculation of VaR in a South African context. A volatility updating method was shown to afford improved results, and the application of EVT at the higher VaR confidence levels gave much better results than the basic RiskMetrics™ assumption of conditional normality. Even the best results were, however, not as good as those obtained in the literature, indicating that some additional work needs to be performed to take into account the unique characteristics of the South African market. One possible approach, involving asymmetric GARCH models was suggested, but time constraints did not allow a full investigation into this matter.

In terms of the practical implementation of the methods discussed in this study, one of the criticisms that the approaches may be subjected to is the large amount of data required. It should be noted, however, that this is an aspect of VaR calculation methods in general, and various strategies have been devised to overcome this problem. These can be applied equally well to the approaches discussed here, with the benefit that EVT is likely to provide a more accurate estimate, as it is likely to make more efficient use of the available data. In fact, since the methods discussed in this study are extensions of existing methods, most strategies devised to handle differing situations (including portfolios consisting of non-linear instruments such as derivatives) can be employed.

There are additional aspects to the EVT-based methods of VaR calculation, such as extending a one day VaR forecast to a multiple day forecast, and *expected shortfall*, a measure of the expected loss, given that the loss exceeds the VaR estimate.[24d] Before an investigation into these topics, in a South African context, can be embarked upon, the basic VaR estimation procedure should be shown to be sound, which has yet to be done.

References

1. Jorion, P., http://www.gsm.uci.edu/~jorion/oc/case.htm
2. Schachter, B, *Financial Engineering News*, no.1, 1997
3. Hull, J.C., *Options, Futures, and Other Derivatives*, 4th Ed., Prentice Hall, 1999, p342
4. J.P. Morgan, *RiskMetrics Technical Manual*, J.P. Morgan Bank New York, 1996, p7
5. Mathworks, *GARCH Toolbox Tutorial*, 2000, p1
6. Hull, J.C., *Options, Futures, and Other Derivatives*, 4th Ed., Prentice Hall, 1999, p355
7. Taleb, N, http://pw1.netcom.com/~ntaleb/jorion.html, 1997
8. Dowd, K, *Financial Engineering News*, no.11, 1999
9. McNeil, A.J., "*Extreme Value Theory for Risk Magagers*", ETH preprint (www.math.ethz.ch/~mcneil/pub_list.htlml), 1999
10. Embrechts, P., Kluppelburg, C., and Mikosch, T., *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin, 1997, p121
11. Embrechts, P., Kluppelburg, C., and Mikosch, T., *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin, 1997, p152
12. Embrechts, P., Kluppelburg, C., and Mikosch, T., *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin, 1997, p128
13. Embrechts, P., Kluppelburg, C., and Mikosch, T., *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin, 1997, p132
14. Embrechts, P., Kluppelburg, C., and Mikosch, T., *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin, 1997, p160
15. Embrechts, P., Kluppelburg, C., and Mikosch, T., *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin, 1997, p165
16. Bensalah, Y., "*Steps in Applying Extreme Value Theory to Finance: A Review*", Bank of Canada Working Paper, www.bankofcanada.ca/en/res/wp00-20.htm, 2000
17. Danielsson, J., and De Vries, C., "*Beyond the Sample: Extreme Quantile and Probabiltiy Estimation*" Timbergen Institute, Erasmus University, Discussion Paper, 1997
18. Mathworks, *GARCH Toolbox Tutorial*, 2000, p2
19. Engle, R.F., *Econometrica*, **50**, 987-1007, 1982
20. Bollerslev, T., *Journal of Econometrics*, **31**, 307-27, 1986

21. Hamilton, J.D., *Time Series Analysis*, Princeton University Press, 1994, p58

22. Hamilton, J.D., *Time Series Analysis*, Princeton University Press, 1994, p59

23. Longin, F.M., "*From Value-at-Risk to Stress Testing: the Extreme Value Approach*", CERESSEC Working paper 97-004, 1997

24. a) Embrechts, P., Resnick, S., and Samorodnitsky, G., "*Extreme Value Theory as a Risk Management Tool*", ETH preprint (www.math.ethz.ch/~embrechts), 1998

b) Danielsson, J., and De Vries, C., "*Value-at-Risk and Extreme Returns*", Timbergen Institute, Erasmus University, Discussion Paper, 1997

c) McNeil, A.J., "*Calculating Quantile Risk Measures for Financial Return Series using Extreme Value Theory*", ETH preprint, 1998

d) McNeil, A.J. and Frey, R., "*Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series*", ETH preprint, 2000

25. Embrechts, P., Kluppelburg, C., and Mikosch, T., *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin, 1997, p331

26. MATLAB is a registered trademark of The Mathworks, Inc.

27. VBA was used within Microsoft's Excel 97

28. Embrechts, P., Kluppelburg, C., and Mikosch, T., *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin, 1997, pp 292-293

29. a) Netter, J., Wasserman, W., and Whitmore, G.A., *Applied Statistics*, 4th ed., Allyn & Bacon, Boston, 1993, p493

b) D' Agostino, B., and Stephens, M.A. ed., *Goodness-of-Fit Techniques*, Marcel Dekker, Inc., 1986

30. Netter, J., Wasserman, W., and Whitmore, G.A., *Applied Statistics*, 4th ed., Allyn & Bacon, Boston, 1993, p937

31. Embrechts, P., Kluppelburg, C., and Mikosch, T., *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin, 1997, p357

32. Borse, G.J., *Numerical Methods with MATLAB®*, PWS Publishing Company, 1997

33. Levy, G., *Financial Engineering News*, no.22, 2001

34. Engle, R.F. and Ng, V.K., *Journal of Finance*, **XLVIII**, no.5, p1749, 1993

35. Glosten, L.R., Jagannathan, R., and Runkle, D.E., *Journal of Finance*, **XLVIII**, no.55, p1779, 1993

*Appendix A

Appendix A contains the MATLAB code used to perform the backtest of the historical simulation method. Only a single script file was required, i.e. no additional function calls were necessary.

```
% Script file to test VaR historical simulation method
load c:\ant\maths\data\returns.prn; % Load returns for the 9 stocks
weights=rand(9,500);
varr=zeros(7,500);
nextday=zeros(7,500);
varcount=zeros(7,500);
averagevar=zeros(1,7);
% Generate 500 random portfolio weights
for i=1:500
  total=0;
  for j=1:9
    total = total + weights(j,i);
  end;
  for k=1:9
    weights(k,i)=weights(k,i)/total;
  end
end
% Generate matrix of returns for the 500 portfolios
portfolioreturns=returns*weights;
for t=1500:2486 % Perform Backtest
   sorted=sort(portfolioreturns(t-1499:t,:));
   varr(1,:)=sorted(75,:);
   varr(2,:)=sorted(38,:);
   varr(3,:)=sorted(15,:);
   varr(4,:)=sorted(8,:);
   varr(5,:)=sorted(4,:);
   varr(6,:)=sorted(2,:);
   varr(7,:)=sorted(1,:);
   for k=1:7
      nextday(k,:)=portfolioreturns(t+1,:);
   end
   varcount=varcount+(nextday<varr);
end
averagevar=(mean(varcount'))
stdev=std(varcount')
```

* Code and stock price data will be gladly supplied via e-mail, if requested. Requests may be sent to symant02@science.uct.ac.za, or to daniel.polakow@cadiz.co.za.

Appendix B

Appendix B contains the MATLAB code used to perform the backtest of Danielsson and De Vries' VaR estimation method. The main script file that executed the backtest is shown, as well as additional functions that were called.

Main Script File

```
% Script file to test Danielsson % De Vries' EVT-based
% VaR calculation method. For each of a specified number
% of random number of portfolios, VaR is calculated
% approximately 1000 times, and each VaR estimate is compared
% to the realised return the following day.


t1=clock;
numportf=500;
numtests=986;
rand('seed',5); % Set seed for random number generator
q=[0.95 0.975 0.99 0.995 0.9975 0.999 0.9995 0.99975 0.9999 0.99995];
q=q';
numvar=length(q);
varcount=zeros(numvar,numportf)
portfcount=0;
testcount=0;

% Set up matrix of stock returns
SP=load('Stock_Prices4Matlab.txt'); % SP is a matrix of available stock prices
                                    % for the 9 stocks
sizeSP=size(SP);
L=sizeSP(1);
returns=(SP(2:L,:)-SP(1:L-1,:))./SP(1:L-1,:); % Matrix of stock returns
clear SP; % Reclaim memory used by SP matrix


% Generate matrix of random portfolio weights
weights=rand(sizeSP(2),numportf);
sumweights=sum(weights);
temp=sumweights;
for i=1:sizeSP(2)-1
   temp=[temp; sumweights];
end
weights=weights./temp;

% Generate matrix of portfolio returns
PR=returns*weights;
PR=-log(1+PR); % Convert to continuously compounded losses
clear returns;  % Reclaim memory

% Loop over portfolios & numtests VaR calculations
```

```
for i=1:numportf
   portfcount=portfcount+1;
   testcount=0;
   for j=1:numtests
      testcount=testcount+1;
      a=j;
      b=a+1499;

      % Create vector of losses to examine in this loop
      losses=PR(a:b,i);

      % Calculate value-at-risk
      Q1=sort(losses);
      n=length(Q1);
      k=round(0.1*n);
      tail=Q1(n-(k-1):n);
      threshold=Q1(n-k);
      alpha=hillest(k,tail,threshold);
      V=DDquantile(q,n,k,alpha,threshold);

      % Compare V to next day's realised return
      nextday=ones(length(V),1)*PR(b+1,i);
      varcount(:,i)=varcount(:,i) + (nextday > V);
      portfcount
      testcount
   end
end
diary c:\ant\ddvarcount.dat % Write results to disc.
averagevar=mean(varcount')
stdev=std(varcount')
t2=clock;
seconds=etime(t2,t1); % Record elapsed time
minutes=seconds/60
hours=minutes/60
varcount'
diary off;
```


Called Functions

```
function alpha=hillest(k,tail,threshold)
% Function which calculates the Hill estimate of the tail index

A=tail/threshold;
B=log(A);
C=sum(B);
alpha=1/(C/k);


function V=DDquantile(q,n,k,alpha,threshold)
% Function to calculate required VaR estimates

A=(n/k)*(1-q);
B=A.^(-1/alpha);
V=B*threshold;
```

Appendix C

Appendix C contains the MATLAB code used to perform the backtest of McNeil and Frey's VaR estimation method. The main script file that executed the backtest is shown, as well as additional functions that were called. "Built in" MATLAB functions, such as those from the GARCH Toolbox are not included.

Main Script File

```
% Script file to test McNeil's EVT-based
% VaR calculation method. For each of a specified number
% of random number of portfolios, VaR is calculated
% approximately 1000 times, and each VaR estimate is compared
% to the realised return the following day.
%
% This file is for an autoregressive component of 0, with
% constant conditional mean.

t1=clock;
numportf=500;
numtests=986;
rand('seed',150); % Set seed for random number generator
complexcount=0;
numvar=10;
varcount=zeros(numvar,numportf);
normvarcount=zeros(numvar,numportf);
q=[0.95 0.975 0.99 0.995 0.9975 0.999 0.9995 0.99975 0.9999 0.99995];
q=q';
nq = norminv(q,0,1);
portfcount=0;
testcount=0;

% Set up matrix of stock returns
SP=load('Stock_Prices4Matlab.txt'); % SP is a matrix of available stock prices
                              % for the 9 stocks
sizeSP=size(SP);
L=sizeSP(1);
returns=(SP(2:L,:)-SP(1:L-1,:))./SP(1:L-1,:); % Matrix of stock returns
clear SP; % Reclaim memory used by SP matrix

% Generate matrix of random portfolio weights
weights=rand(sizeSP(2),numportf);
sumweights=sum(weights);
temp=sumweights;
for i=1:sizeSP(2)-1
   temp=[temp; sumweights];
end
weights=weights./temp;

% Generate matrix of portfolio returns
```

```
PR=returns*weights;
PR=-log(1+PR); % Convert to continuously compounded losses
clear returns;  % Reclaim memory

% Loop over portfolios & numtests VaR calculations
for i=1:numportf
   portfcount=portfcount+1;
   testcount=0;
   for j=1:numtests
      testcount=testcount+1;
      a=j;
      b=a+1499;

      % Create vector of losses to examine in this loop
      losses=PR(a:b,i);

      spec=garchset('R',0,'M',0,'P',1,'Q',1,'TolFun',1e-5); % Specify ARMA model
      % Pass losses vector to GARCH toolbox to estimate parameters
      [coeff, errors, LLF, innovations, sigma]=garchfit(spec,losses);

      % Calculate realisation of i.i.d noise process
      residuals=innovations./sigma;

      % Fit GPD to tail of residuals
      residuals=sort(residuals);
      n=length(residuals);
      k=round(0.1*n);
      tail=residuals(n-(k-1):n);
      threshold=residuals(n-k);
      Y=tail-threshold;
      fnc='likely2';
      deriv='dfdx';
      r=newton(-0.05,fnc,deriv,Y,0.0001);
      % Check for r complex
      complexcount = complexcount + not(r==r');
      e=ehat(r,Y); % Calculate shape parameter
      beta=-(e/r);

      % Calculate one step ahead forecast of conditional standard deviation
      [sFcast, yFcast]=garchpred(coeff, losses);

      % Calculate value-at-risk
      zq=noisequantile(threshold,e,beta,k,n,q);
      C=garchget(coeff,'C');
      V=varpred(C,sFcast,zq);

      % Compare V to next day's realised return
      nextday=ones(length(V),1)*PR(b+1,i);
      varcount(:,i)=varcount(:,i) + (nextday > V);
      portfcount
      testcount

      % Calculate VaR assuming normally distributed innovations Z
      Vnorm=varpred(C,sFcast,nq);
      normvarcount(:,i)=normvarcount(:,i) + (nextday > Vnorm);
   end
```

```
end
diary c:\varcount.dat
averagevar=mean(varcount')
stdev=std(varcount')
t2=clock;
seconds=etime(t2,t1); % Record elapsed time
minutes=seconds/60
hours=minutes/60
varcount'
diary off;
```

Called Functions

```
function h_t=likely(t,Y)
 % Function used to calculate GPD parameters

    A=1-t*Y;
    D=log(A);
    ehat=sum(D)/length(Y);
    F=Y./A;
    G=sum(F);
    h_t=1/t + ((1/ehat +1)*G)/length(Y);


function deriv=dfdx(t,Y)
% Function to calculate derivative of
% function "likely2"

A1=1-t*Y;
A2=Y./A1;
A=sum(A2);
B1=log(A1);
B=sum(B1);
term1=(A/B)*(A/B);
term2=1/B + 1/length(Y);
C1=Y.^2;
C2=A1.^2;
C3=C1./C2;
C=sum(C3);
deriv=-(1/t)*(1/t) + term1 + term2*C;
```

```
function r = Newton(x,fnc,dfdx,Y,small)
%Newton finds the root of the dummy function fnc starting from the
%  initial guess x.  The analytic derivative of this function must
%  also be coded as an m-file and its name transferred in dfdx.
%  Convergence is attained when |f1-f2| < small.  For example,
%  fnc='sin', dfdx='cos', small=1.e-3.
%==============================================================
 if isstr(fnc)~=1 | isstr(dfdx)~=1;
   error('Both fnc and dfdx must be strings');
 end
%==============================================================
 r = x;             dx = 1;        i = 0;
 while abs(dx) > small;
   i = i + 1;
   if i>20; error('excessive iterations (iter = 20)');end
   df = feval(dfdx,r,Y);
   if abs(df) < small
           disp('small derivative, method diverging')
        disp('to continue hit any key')
           disp('to interrupt, hit CTL-C')
   end
   df = feval(dfdx,r,Y);
   if abs(df)<small; error('small derivative, method diverging');end
   dx = -feval(fnc,r,Y)/df;
   r = r + dx;
 end


function e=ehat(r,Y)
% Function which calculates the maximum likelihood estimate
% of the shape parameter, given the appropriate value of tau.

A=1-r*Y;
B=log(A);
e=(1/length(Y))*sum(B);



function zq=noisequantile(u,e,beta,k,n,q)
% Function which calculates the quantiles of the noise
% distribution function for given probabilities q.
A=1-q;
B=(n/k)*A;
C=(B.^(-e)) - 1;
D=(beta/e)*C;
zq=u+D;


function V=varpred(C,sigtplus1,zq)
% Function to calculate value at risk using
% parameters from the predictive distribution
% function of portfolio returns
% This is for zero autoregression component in the ARMA specification
V = C + sigtplus1*zq;
```

Appendix D

Appendix D contains the MATLAB code used to perform the backtest of the GJR-GARCH modification of McNeil and Frey's VaR estimation method. In addition to the main script file, the modified MATLAB Toolbox functions are shown. Called functions that are identical to those displayed in Appendix C are not shown.

Main Script File

```
% Script file to test GJR-GARCH
% modification of McNeil's EVT-based
% VaR calculation method. For each of a specified number
% of random number of portfolios, VaR is calculated
% approximately 1000 times, and each VaR estimate is compared
% to the realised return the following day. This file is
% based on the trial script file mactrial.m
% This file is for an autoregressive component of 0, with
% constant conditional mean.

path(path,'C:\matlabr12_2\work\asymgarch');
t1=clock;
numportf=5;
numtests=986;
rand('seed',10531918); % Set seed for random number generator
complexcount=0;
numvar=10;
varcount=zeros(numvar,numportf);
normvarcount=zeros(numvar,numportf);
q=[0.95 0.975 0.99 0.995 0.9975 0.999 0.9995 0.99975 0.9999 0.99995];
q=q';
nq = norminv(q,0,1);
portfcount=0;
testcount=0;

% Set up matrix of stock returns
SP=load('Stock_Prices4Matlab.txt'); % SP is a matrix of available stock prices
                              % for the 9 stocks
sizeSP=size(SP);
L=sizeSP(1);
returns=(SP(2:L,:)-SP(1:L-1,:))./SP(1:L-1,:); % Matrix of stock returns
clear SP; % Reclaim memory used by SP matrix

% Create matrix to store one-step ahead predicted volatilities
predvol=zeros(L,numportf);

% Generate matrix of random portfolio weights
weights=rand(sizeSP(2),numportf);
sumweights=sum(weights);
temp=sumweights;
```

```
for i=1:sizeSP(2)-1
   temp=[temp; sumweights];
end
weights=weights./temp;

% Generate matrix of portfolio returns
PR=returns*weights;
PR=-log(1+PR); % Convert to continuously compounded losses
clear returns;  % Reclaim memory

% Loop over portfolios & numtests VaR calculations
for i=1:numportf
   portfcount=portfcount+1;
   testcount=0;
   for j=1:numtests
      testcount=testcount+1;
      a=j;
      b=a+1499;

      % Create vector of losses to examine in this loop
      losses=PR(a:b,i);

      % Fit asymmetric GJR model
      [kappa , alpha , beta ,gamms] = agarch(losses , 1 , 1);
      [cvFcast,cvSeries] = agarchpred(losses , kappa , alpha , beta , gamms, 1);
      sFcast=sqrt(cvFcast);



      % Calculate realisation of i.i.d noise process
      aresiduals=losses./sqrt(cvSeries);

      % Fit GPD to tail of residuals
      residuals=sort(aresiduals);
      n=length(residuals);
      k=round(0.1*n);
      tail=residuals(n-(k-1):n);
      threshold=residuals(n-k);
      Y=tail-threshold;
      fnc='likely2';
      deriv='dfdx';
      r=newton(-0.05,fnc,deriv,Y,0.0001);
      % Check for r complex
      complexcount = complexcount + not(r==r');
      e=ehat(r,Y); % Calculate shape parameter
      beta=-(e/r);



      % Calculate value-at-risk
      zq=noisequantile(threshold,e,beta,k,n,q);

      V=varpred(0,sFcast,zq);

      % Compare V to next day's realised return
      nextday=ones(length(V),1)*PR(b+1,i);
```

```
        varcount(:,i)=varcount(:,i) + (nextday > V);
        portfcount
        testcount

        % Calculate VaR assuming normally distributed innovations Z
        Vnorm=varpred(0,sFcast,nq);
        normvarcount(:,i)=normvarcount(:,i) + (nextday > Vnorm);

        % Store predicted volatility
        predvol(j,i)=sFcast;
    end
end
diary c:\varcount.dat
averagevar=mean(varcount')
stdev=std(varcount')
t2=clock;
seconds=etime(t2,t1);
minutes=seconds/60
hours=minutes/60
varcount'
diary off;
```

Modified Toolbox Functions

```
function [kappa , alpha , beta, gamms] = agarch(u , p , q , options)
% Modification of MATLAB GARCH parameter estimation procedure.
%
% [Kappa , Alpha , Beta, gamms] = agarch(U , P , Q)
%
%


%
% Error-checking on P and Q. Note that I allow P to be zero, with
% the understanding that a GARCH(0,Q) process is an ARCH(Q) process.
%

if (length(p) > 1) | any(p <  0)
  error(' P must be a non-negative scalar.')
elseif isempty(p)
  error(' P is an empty matrix.')
end

if (length(q) > 1) | any(q <= 0)
  error(' Q must be a positive scalar.')
elseif isempty(q)
  error(' Q is an empty matrix.')
end

%
% Ensure we have a single, univariate time series COLUMN vector.
%

if size(u,2) > 1
  error(' Innovations time series U must be a single column vector.')
elseif isempty(u)
```

```
  error(' Innovations time series U is empty.')
end

%
% GARCH(P,Q) processing requires pre-sample values for conditioning. That is, we
% require P pre-sample values of h(t), the conditional variance of the residuals,
% or innovations, u(t), and Q pre-sample values of u(t) itself to 'jump-start' the
% process. To be safe, allocate M = max(P,Q) pre-sample lags for both h(t) and u(t).
%

m  =  max(p,q);   % # of pre-sample lags needed to 'jump-start' the process.

%
% Make an initial guess for the GARCH parameter vector to be estimated. The vector
% is initialized based on a (conservative) application of the following GARCH
% constraints:
%
%   (1) k > 0
%   (2) a(1),a(2),...a(P) , b(1),b(2),...b(Q) >= 0
%   (3) a(1) + a(2) + ... + a(P) + b(1) + b(2) +... + b(Q) <= 1
%
% The initial 'guess' is based on setting the last constraint equal to 0.5.
%

guess  =  1/(4*m);
alpha  =  guess(ones(p,1));
beta   =  guess(ones(q,1));
kappa  =  guess;
gamms  =  guess(ones(q,1));

%
% Perform constrained optimization.
%

lowerBounds        =  zeros(4, 1);
lowerBounds(3)=0;                        % Linear inequalities captured
lowerBounds(4)=0;
upperBounds        =  [];                   % as lower bounds constraints.

gradientFunction    =  [];

summationConstraintA =  [0  ones(1,p)  ones(1,q) 0 ];  % Linear inequality related to
summationConstraintB =  1;                          % covariance-stationary constraint.

%
%  Allow a user-specified OPTIONS structure if one exists, but set default
%  to 'iterative display' with 'diagnostics' printed to the screen.
%

if (nargin <= 3) | isempty(options)
  options  =  optimset('fmincon');
  options  =  optimset(options , 'Display'     , 'iter');
  options  =  optimset(options , 'Diagnostics' , 'on');
  options  =  optimset(options , 'LargeScale'  , 'off');
end
```

```
%
% Set linear inequality of the covariance-stationarity constraint of
% the conditional variance, Ax <= b. Since the conditional variance
% (ALPHA(i) for i = 1,2,...P, BETA(j) for j = 1,2,...Q) parameters
% are constrained to be non-negative, the covariance-stationarity
% constraint is just a summation constraint. Also, adjust the
% summation constraint to reflect a tolerance offset from a fully
% integrated conditional variance condition.
%

summationConstraintB = summationConstraintB - 2*optimget(options, 'TolCon', 1e-6);

%
% Estimate the parameters.
%

[garchParameters , LLF    , ...
 EXITFLAG        , OUTPUT , ...
 LAMBDA  , GRAD  , HESSIAN] =  fmincon('agarchllf'          , [kappa ; alpha ; beta ; gamms] , ...
                               summationConstraintA  , summationConstraintB  , ...
                               [] , [] , lowerBounds , upperBounds          , ...
                               [] , options          , u , p , q);
%
% Over-write all GARCH constraint-violating parameters that are less than
% zero. This will, occasionally, occur because FMINCON may violate constraints
% ever so slightly. When a constraint violation does occur, it is EXTREMELY
% small (e.g. -3.51422153312504e-013 is very nearly 0).
%

garchParameters(find(garchParameters    <  0)) = 0;          % All coefficients
garchParameters(find(garchParameters(1) <= 0)) = realmin;    % Constant term Kappa.

%
% Assign (decode) the estimated GARCH(P,Q) coefficients to the output parameters.
%

kappa  =  garchParameters(     1);
alpha  =  garchParameters( 2);
beta   =  garchParameters( 3);
gamms   =  garchParameters( 4);
garchParameters
```

```
function LLF = agarchllf(parameters , u , p , q)
%Modification of MATLAB Toolbox GARCH log-likelihood function
%to allow use of GJR-GARCH
%
%


%
% Over-write all parameters less than or equal to zero to
% prevent the LLF from becoming minus infinity or complex.
%

parameters(find(parameters <= 0)) = realmin;


m  =  max(p,q);   % # of pre-sample lags needed to 'jump-start' the process.

stdEstimate =  std(u,1);                    % Estimate the unconditional STD of u(t).
u          =  [stdEstimate(ones(m,1)) ; u];  % Prepend M samples to u(t).
T          =  size(u,1);                  % The total (original + M padded samples) length of u(t).

%
% Since M samples of the unconditional standard deviation of u(t) have been
% assigned to the first M samples of u(t), assign u^2(1) to the first M
% samples to h(t) to 'jump-start' the process.
%

h  =  u(1).^2;
h  =  [h(ones(m,1)) ; zeros(T-m,1)];   % Pre-allocate the h(t) vector.

%
% Form the h(t) time series and evaluate the log-likelihood function.
% Note that LLF estimation is based on the original u(t) innovations
% sequence, and ignores the M pre-pended samples.
%
indicator = (u > 0);

for t = (m + 1):T
   h(t) = parameters' * [1 ; h(t-(1:p)) ; u(t-(1:q)).^2 ; (u(t-(1:q)).^2)*indicator(t-(1:q))];
end

t    = (m + 1):T;
LLF  =  sum(log(h(t))) + sum((u(t).^2)./h(t));
LLF  =  0.5 * (LLF  +  (T - m)*log(2*pi));
```

[52] MacMillan, L.W. "An Analytic Approximation for the American Put Price", Advances in Futures and Options Research. 1986

[53] Jaillet, P., Lamberton, D. and Lapeyre, B. "Variational Inequalities and the Pricing of American Options", Acta Applicandae Mathematicae. 1990.

[54] B. Thomas. Something to shout about. RISK, 6:56{58, May 1993.