The Best of Two Worlds—Combining Conditional Volatility Models with Extreme Value Theory to Calculate Value at Risk

---

An empirical evaluation of VaR estimation approaches

Bachelor's Thesis

Bachelor Programme in Business and Economics

Department of Economics

Lund, January 17, 2012

LUND UNIVERSITY

School of Economics and Management

Authors:
Benjamin Kjellson
Simon Koskinen Rosemarin

Supervisor:
Hans Byström

Abstract

We compare Value at Risk estimates from an AR(1)-GARCH(1,1) model with t- or normally distributed innovations, to estimates from an AR(1)-GARCH(1,1) model where the Peak-Over-Threshold method is applied to the tails of the innovations. Using the Christoffersen backtest, we find that the performance of the second type of model is superior to the first, particularly at high confidence levels for the VaR estimate.

Acknowledgements

We would like to thank our advisor Hans Byström for his support and often lightning fast responses to queries. We are also grateful for the education we have received at the Center for Mathematical Sciences, Lund University, without which this essay would have been too challenging an undertaking.

Caveat Lector

Because of our particular educational backgrounds[1], we intentionally chose a thesis topic that would challenge the skills we have attained so far in each of our degrees. Therefore, some mathematical and statistical literacy is expected of the reader. However, a quick search on Google or Wikipedia can often fill a knowledge gap just enough to continue reading, and we encourage the reader to perform such a search whenever needed.

[1] Aside from the Bachelor's program in Business and Economics, Benjamin is a student of the BSc. programme in Mathematics at the Faculty of Science, Lund University, specializing in Mathematical Statistics; and Simon is a student of the MSc. programme in Engineering Mathematics at the Faculty of Engineering, Lund University, specializing in Financial Modeling.

Contents

Page

List of Tables

List of Figures

1 Introduction

1.1 Background

Risk management is “the process by which various risk exposures are identified, measured, and controlled” (Jorion 2001, p. 3). The financial crisis that started in 2007 has made obvious the need for risk management in finance, and has illuminated some of the weaknesses of currently used methods of controlling risk. Even so, the theory behind financial risk management has advanced greatly since the groundbreaking work of Harry Markowitz in the 1950s, and has taken its place as a distinct subfield of financial theory (Dowd 2005, p. 1). Financial risk management has evolved in parallel with financial markets and instruments therein. Growth in trading activity in traditional markets; the introduction of new financial instruments; new, emerging markets; and the increased interdependence of these, have meant new challenges to financial risk management (Dowd 2005, p. 3). These developments have brought with them an increase in volatility, and several incidents involving misuse of derivatives and leverage has increased general awareness of risk, an awareness that has stimulated the development of risk management practises (Dowd 2005, p. 4). Value at Risk (VaR), which can be loosely defined as an estimate of the worst loss over a given time period with a specified confidence level, was developed to be an easily comprehensible method of quantifying financial market risk. VaR has become the industry standard for measures of financial market risk, but there are no set rules for how it is to be implemented. As such, the question of which model works best remains open. While no definite answer may exist, some models work better than others, and in this essay, we intend to compare some of the traditionally used models to ones that are of more recent development.

1.2 Problem Discussion & Purpose

Traditional parametric approaches to calculating Value at Risk have often either ignored the fat-tailedness of financial distributions (e.g. assumed normality), or ignored volatility clustering, or both. In this essay, we intend to combine tools for modelling conditional volatility from time series analysis with the statistical theory of extreme values, to produce a model for Value at Risk that rests on more realistic assumptions. We thus ask: can the combination of conditional heteroscedastic models with extreme value theory distributions produce better VaR estimates than conditional heteroscedastic models alone? We test this empirically on nine data sets, using an AR(1)-GARCH(1,1) model with either t-or normally distributed innovations as a base, and use the Peak-Over-Threshold model on the tails of the residuals in the EVT case.

1.3 Delimitations

Given the time constraints involved, and our realism regarding the range of issues a Bachelor's thesis is expected to cover, it is neither possible nor desirable to cover all aspects of Value at Risk, much less other risk measures and financial risk management overall. Therefore, we have limited our research to cover only certain aspects of the VaR estimation process and related issues.

First, Value at Risk is, despite its flaws, still the main tool within risk management for quantifying financial market risk. The recommendations on banking laws and regulations, Basel I and Basel II, state that the preferred approach for quantifying market risk is VaR, and in the words of Dowd (2005, preface p. xiv), "there will be a need to estimate VaR for as long as there is a demand for VaR itself: if someone wants the number, then someone will want to estimate it, and whether anyone should want the number in the first place is another matter". For these reasons, we limit ourselves to testing Value at Risk. However, we do provide a brief discussion of alternative risk measures in section 2.1.4, covering among other things Expected Shortfall (ES), a risk measure that has some advantages over VaR, but has yet to see the same widespread use.

Second, it was necessary to limit the number of approaches to VaR estimation. Even considering just the methods that take account of the stochastic nature of volatility, there are a wealth of models to choose from. We have therefore taken the pragmatic approach of choosing models that have most support in financial journals and books. We include the historical simulation approach in our study mainly as a reference measure, but leave out other common approaches to VaR estimation, such as Monte Carlo simulations.

Third and last, we have restricted the analysis to three asset classes: stock indices, individual stocks, and commodities (oil, gold, wheat). These asset classes are of interest since they may be representative of (parts of) portfolios held by institutional investors, so that accurate VaR-models for these may be of actual use. Historically, these asset classes have also shown definite signs of fat-tailedness in their return distributions, motivating the use of models that can account for large deviations, such as models from extreme value theory. We have tested our models on nine data sets in total, each covering roughly the last 30 years or so of daily observations.

1.4 Previous research

We now present some previous research specifically on the use of Extreme Value Theory to calculate Value at Risk. We do not cover previous research of extreme value distributions in finance in general, nor do we cover conditional models in finance; the number of such models and the number of studies using them is too large to cover well. However, we would like to mention the paper by Engle (2001) as a good introduction to the uses of ARCH/GARCH models and their extensions. The background and theory underlying the models used in this essay are presented in section 2, but we would like to take the opportunity now to direct the interested reader to some litterature that we have found useful. For Value at Risk, Jorion (2001) and Dowd (2005) are perhaps the two most well-known and cited works, but we have also made heavy use of Alexander (2008) and, particularly for mathematical depth, the book by McNeil et al. (2005). For time series analysis, we have found the book by Tsay (2010) very useful. Regarding Extreme Value Theory in finance, we have seen Embrechts et al. (1997) cited in almost every article on EVT and VaR we've read. To the reader wishing to learn more about EVT in general, we recommend Coles (2001) as a good place to start. Lastly, the reader interested in the presence of fat tails in financial data may wish to start with the classical studies of Mandelbrot (1963) and Fama (1965).

Most mentions of EVT and VaR together can be traced to the late 1990s; for example, the book by Embrechts et al. (1997) provides several examples of VaR-calculations by use of extreme value distributions. Daníelsson & de Vries (2000) study unconditional EVT as applied to VaR, and make a comparison to the conditional approach. They suggest that the choice of methodology should depend on the purpose of estimating VaR, stating that an unconditional approach can yield more stable estimates and provide a better guard against disasters, while the conditional approach may protect the portfolio better in the period *after* the initial jump in volatility. Following their paper, there were a number of studies applying unconditional EVT models to calculate VaR; see for example Bekiros & Georgoutsos (2005), Gilli & Këllezi (2006), or Tolikas et al. (2007). On the opposite side, the paper by McNeil & Frey (2000) was the first to calculate Value at Risk by use of conditional EVT; first fitting an AR(1)-GARCH(1,1) model to the data and then applying the POT-model to the standardized residuals. The paper by McNeil & Frey was very influential, and has been the inspiration of many papers on conditional EVT and VaR, this one included. Other examples are the papers by Byström (2004), Kuester et al. (2006), and Ozun et al. (2010).

2 Theoretical background

2.1 Value at Risk

2.1.1 Defining Value at Risk

Value at Risk (VaR) is a risk measure with origins in the internal risk models developed by banks in the late 1970s and 1980s. Following developments at J.P. Morgan under the direction of Dennis Weatherstone, this method had by the mid-1990s established itself as the dominant measure of financial risk in the industry (Dowd 2005, p. 11). Given a time horizon and a confidence level, the Value at Risk is an estimate of the maximal loss we may incur under normal market conditions. For example, VaR can tell us what our worst loss is expected to be 99 days out of 100. Equivalently, VaR gives us an estimate of the minimal loss associated with extraordinary market conditions. Regardless, both definitions imply the same method of quantifying financial risk (Tsay 2010, pp. 325-326). From a mathematical point of view, (percentage) Value at Risk can be defined via the distribution of the portfolio returns in the following way:

Definition. *Given some confidence level $q \in (0,1)$ and a risk horizon of h days, the h-day value at risk, expressed as a percentage of the portfolio value at time t, is given by the smallest number x_q such that the probability that the loss X_t exceeds x_q is no larger than $\alpha = (1-q)$. Formally,*

$$VaR_{q,h} = \inf\{x_q \in \mathbb{R} : P(X_{t;h} > x_q) \leq 1-q\} = \inf\{x_q \in \mathbb{R} : F_{X;h}(x_q) \geq q\},$$

where $F_{X;h}$ is the distribution function of the losses over a h-day period.

The perhaps more common way to express VaR is in terms of an actual money amount—this is achieved by simply multiplying the percentage VaR defined above by the value of the portfolio P_t. In this essay, we use the percentage VaR definition since it is more convenient to deal with the loss distribution than with the value of some fictional portfolio of an arbitrary size. We will also take h as given and omit its inclusion in the notation, to avoid cluttering. Note that the losses are simply the negative returns (e.g. simple or compounded); this is done so that the VaR will be a positive number. It is important to note that the VaR does not contain any information regarding the size of the losses beyond the point x_q (McNeil et al. 2005, p. 38); this is in fact one of the arguments against the use of VaR as a risk measure. More on this later.

Probabilistically, VaR is a quantile of the loss distribution, with q usually set to 0.95 or 0.99; the confidence level for the VaR prediction depends mainly on the approach to risk the user has. In general, the more conservative the user, the higher the value of q chosen (Alexander 2008, pp. 13-14). Typical values for the time horizon within market risk management are 1 or 10 days (McNeil et al. 2005, p. 38). For example, the risk horizon under Basel II banking regulations for VaR is set to 10 days, and this framework also requires VaR to be derived at the 99 % confidence level (Dowd 2005, p. 52). These regulations further state that the 10-day VaR must be estimated on a daily basis, but they also allow banks to compute the 10-day VaR by multiplying the 1-day VaR by $\sqrt{10}$ (the so-called square root of time rule). The Basel agreement also requires (with a few exceptions) that the historical data sample used to estimate VaR to cover at least 1 year, but there are no restrictions regarding type of model used for the actual estimation (Alexander 2008, p. 408).

2.1.2 Advantages of VaR

VaR has a number of advantages, that has contributed to making it superior to earlier models. Dowd (2005, p. 12) lists a few of the attractions of the VaR method:

- VaR is simple and easy to understand: when measured in money terms, it is easy to interpret and convey to others.
- VaR is probabilistic, giving the user the probability associated with a certain loss. This sets VaR apart from other risk measures such as duration, convexity, and the Greeks, which only provide numbers without associated probabilities.
- VaR is holistic, both in the sense that it takes all risk factors into account simultaneously, and in that it focuses on the firm's portfolio as a whole, rather than its parts.
- VaR enables aggregation of the risk of individual positions into a measure of portfolio risk, and therefore incorporates the interaction of different risk factors with one another.
- VaR can be applied to all types of portfolios (and even to non-financial assets), and allows for comparisons between the risks of different portfolios.

Dowd (2005, pp. 12-13) goes on to list some of the ways in which VaR can be used:

- VaR can be used by senior management to set an overall target for firm risk, and from this target look at how the risk should be distributed in different parts of the firm.
- VaR can, as proposed in the Basel framework, be used to determine both firmwide and business-unit capital requirements.
- Var can be a useful statistic to include in reports and disclosures of financial positions; it is becoming increasingly popular to include VaR figures in annual reports.
- VaR can be used as a factor in or as a base of decision procedures for investment strategies, hedging, and portfolio management.
- VaR can be used in the design of renumeration schemes for traders and other risk takers in a firm, to make rewards reflect both profits generated *and* risks taken.
- Lastly, VaR can be used to manage more than just financial risk: liquidity risks, credit risks, and operational risks may also be measured by VaR, and together these may help in forming a more complete risk management system for the firm.

2.1.3 Disadvantages of VaR

VaR as a risk measure is not without its faults. In fact, it is no secret that VaR has taken heavy criticism over the last two decades, particularly so in the years after the financial crisis. One of the most vocal opponents of VaR is author and former derivatives trader Nassim Taleb, who has brought to public attention the deficiencies of VaR, and current risk management practises in general, in bestselling books such as *Fooled by Randomness* and *The Black Swan*, as well as in newspaper columns and interviews. In light of this criticism, it may be appropriate to ask: what are the downsides of Value at Risk? One problem lies in the implementation of VaR: different models may generate vastly different estimates, and even theoretically similar models may sometimes produce results that are inordinately different if they have been implemented differently (Dowd 2005, p. 13). Since VaR is computed using the predictive distribution of the returns of the asset or portfolio, one would hope that the model will take into account the parameter uncertainty that is inherent in such a distribution. However, most available methods for calculating VaR fail to do this, because of the difficulties in specifying the predictive distribution (Tsay 2010, p. 328).

There are also principal-agent problems associated with VaR when remuneration schemes are designed to take into account the VaR-constraints imposed on traders and other risk takers. These constraints may incentivize the involved parties to take risks that they know are different on paper than in reality (Dowd 2005, p. 13). For example, if the VaR measure fails to take account of the responses of others when the market moves, then the risk may be understated. The trader can thus take advantage of this by taking positions that are larger than would be allowed if the risk estimate was better. Likewise, a trader could under normal market conditions take positions in which the risk lies in low-probability (but high impact) events, expecting current market conditions to remain stable.

There are also mathematical and statistical reasons as to why VaR may not be a good risk measure. In a highly influential paper, Artzner et al. (1999) proposed a few properties that a “coherent” risk measure should have. Amongst these is the property of subadditivity, essentially saying that for a real valued function $f : A \to B$, f should be such that $f(x + y) \leq f(x) + f(y)\ \forall x, y \in A$. Value at risk does not satisfy this property, since the VaR of a portfolio may in some cases be larger than the sum of the VaRs of the parts that in aggregate make up the portfolio, and likewise for a firm and its business units (McNeil et al. 2005, p. 40). This would in turn imply that there are sometimes no diversification benefits from merging separate portfolios, which contradicts much (if not all) of portfolio theory. However, Daníelsson et al. (2005, p. 12) finds that for for heavy tailed distributions (as found in finance), Value at Risk is “subadditive in the tails, at probabilities that are most relevant for practical applications”. Daníelsson et al. (2005, p. 3) also notes that the subadditivity property may not hold for distributions that are extremely fat tailed (e.g. lacking first moment), or in general for the interior of distributions (but these are in any case of lesser importance in risk management).

Some dangers are also found in the interpretation of VaR, due to estimation error and model risk; the latter defined by McNeil et al. (2005, pp. 40-41) as “the risk that a financial institution incurs losses because its risk-management models are misspecified or because some of the assumptions underlying these models are not met in practice”. McNeil et al. (2005, p. 41) go on to state that these problems become even more apparent when estimating VaR at higher confidence levels (e.g. $q = 0.9995$), and the authors also mention the failure of VaR to properly account for liquidation issues.

2.1.4 Alternatives to VaR

Perhaps the most common criticism of Value at Risk is that it fails to inform us of what happens in cases where losses exceed the VaR estimate. What does the tail of the loss distribution beyond the VaR-quantile look like? A risk measure that can answer this question is *Expected Shortfall* (ES), also known as Conditional Value at Risk. Expected shortfall is the average of the worst $\alpha = 1 - q$ losses, defined mathematically as

$$\begin{aligned} \mathrm{ES}_q &= \frac{1}{1-q}\int_q^1 x_{p;F}\,dp \\ &= \frac{1}{1-q}\int_q^1 \mathrm{VaR}_p\,dp \end{aligned}$$

where $x_{p;F}$ is the generalized inverse, or quantile function, defined by $x_{p;F} = \inf\{x \in \mathbb{R} : F(x) \geq p\}$ (McNeil et al. 2005, pp. 39, 44). While ES is a good (or "coherent") risk measure according to the previously mentioned criteria of Artzner et al. (1999), it may see more use in academia than in actual practical implementations (Einhorn & Brown 2008, p. 20). One reason why this may be true has to do with backtesting. Yamai & Yoshiba (2002) showed that, to achieve the same accuracy in backtesting, ES requires a larger sample than does VaR. Daníelsson (2011, p. 162) adds that, since backtesting ES necessarily involves first estimating VaR, the inaccuracy of the ES is "necessarily much higher than that in a VaR backtest", because the errors in the VaR estimate are carried over to the ES estimate.

Examples of other classes of measures of financial risk are the *spectral* risk measures, *distortion* risk measures, and the measures defined through the mean-variance framework. For more information on these, see chapter 2 of Dowd (2005).

2.2 Time series analysis

We now move on to linear and non-linear time series analysis, and econometric techniques for modelling asset returns. We cover some basic theory and quickly move on to the specific models used in this essay. We draw upon Tsay (2010), chapters 2-3, and McNeil et al. (2005), chapter 4, for this section. In the following, we assume $t \in \mathbb{Z}$ and omit its mention in the notation.

2.2.1 Basic concepts of time series analysis

A time series $\{X_t\}$ is said to be strictly stationary if the joint distribution of $(X_{t_1}, \ldots, X_{t_k})$ is invariant under time shift, i.e. the distribution of $(X_{t_1}, \ldots, X_{t_k})$ is identical to that of $(X_{t_1+\ell}, \ldots, X_{t_k+\ell})$, for $\ell \in \mathbb{Z}$ (Tsay 2010, p. 30). Since this condition is very hard to verify empirically for financial series, a less stringent form of stationarity is instead assumed. A time series $\{X_t\}$ is said to be weakly stationary if both the mean of X_t and the covariance between X_t and $X_{t-\ell}$, $\ell \in \mathbb{Z}$, are invariant under time shift. Specifically, this means that $\mathrm{E}(X_t) = \mu$ and $\mathrm{Cov}(X_t, X_{t-\ell}) = \gamma_\ell$ (the lag-ℓ autocovariance), and from this it can be seen that the first two moments of X_t must be finite if weak stationarity is to hold. Furthermore, we have that $\gamma_0 = \mathrm{Var}(X_t)$ and $\gamma_{-\ell} = \gamma_\ell$, and we can thus define the autocorrelation function (ACF), for the correlation between X_t and $X_{t-\ell}$, as

$$\rho_\ell = \frac{\gamma_\ell}{\gamma_0}, \tag{2.1}$$

so that $\rho_0 = 1$ and $-1 \leq \rho_\ell \leq 1$ (Tsay 2010, pp. 30-31). In section 4, we will use the ACF on one of the data sets used in to estimate VaR later on, to test for and compare the presence of autocorrelation in the raw asset returns and in the residuals generated by fitting an econometric model to the same data.

A time series is called linear if it can be written as

$$X_t = \mu + \sum_{i=0}^{\infty} \psi_i Z_{t-1}, \tag{2.2}$$

with weights ψ_i such that $\psi_0 = 1$, and where $\{Z_t\}$ is a white noise series, i.e. a sequence of independent and identically distributed (i.i.d.) random variables with mean zero and finite variance σ_Z^2 (Tsay 2010, p. 36). Z_t can be thought of as new information arriving at time t, and for this reason it is often called the *innovation* at time t (Tsay 2010, p. 36). Since Z_t is i.i.d. with mean zero, $\mathrm{E}(X_t) = \mu$, and if $\{X_t\}$ is weakly stationary, the series of squared weights which together with σ_Z^2 determine the variance of X_t must be convergent (Tsay 2010, p. 36). This in turn imples that the sequence $\{\psi_i^2\}_{i=0}^{\infty}$ converges to zero. From this follows (see Tsay 2010, p. 37) that the linear dependence of X_t on $X_{t-\ell}$ (e.g. past returns) goes to zero for large ℓ, i.e. $\lim_{\ell\to\infty} \rho_\ell = 0$. With this background, we can now describe the family of models from which those used in this essay have been drawn.

2.2.2 Autoregressive models

A model that describes the linear dependence of X_t on past observations is the autoregressive model of order p, or AR(p)-model:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t, \tag{2.3}$$

with $p \in \mathbb{N}$ indicating that the past p variables $X_{t-i}, i = 1, \ldots, p$, "jointly determine the conditional expectation of $[X_t]$ given the past data" (Tsay 2010, p. 38). Z_t is here a white noise series with mean zero and variance σ_Z^2 as before.

We illustrate some of the properties of autoregressive models by looking at the simplest of such models, namely the AR(1) model. The AR(1) model can be written as:

$$X_t = \phi_0 + \phi_1 X_{t-1} + Z_t, \tag{2.4}$$

Conditional on X_{t-1}, we have that

$$\begin{aligned} \mathrm{E}(X_t | X_{t-1}) &= \phi_0 + \phi_1 X_{t-1}, \text{ and} && (2.5) \\ \mathrm{Var}(X_t | X_{t-1}) &= \mathrm{Var}(Z_t) = \sigma_Z^2. && (2.6) \end{aligned}$$

In essence, this model says that, conditional on X_{t-1}, X_t is not correlated with X_{t-i} for $i > 1$ (Tsay 2010, p. 37). Furthermore, if weak stationarity is assumed, it can be shown that

$$\begin{aligned} \mathrm{E}(X_t) &= \mu = \frac{\phi_0}{1 - \phi_1} && (2.7) \\ \mathrm{Var}(X_t) &= \frac{\sigma_Z^2}{1 - \phi_1^2} && (2.8) \end{aligned}$$

where $\phi_1^2 < 1$. In the proof of the second equality, Tsay (2010, p. 39) in fact shows that the AR(1) model is weakly stationary *if and only if* $|\phi_1| < 1$. Tsay (2010, pp. 39-40) goes on to show that for a weakly stationary AR(1) process, the ACF will decay exponentially with rate ϕ_1 from a starting value $\rho_0 = 1$.

The AR(p) model may be used to provide forecasts of $X_{h+\ell}$, where h is the current time index (the forecast origin) and $\ell \geq 1$ is the number of time steps ahead we wish to forecast (the forecast horizon). It can be shown (see Tsay 2010, pp. 54-56) that the 1-step ahead forecast, given all available information up to and including time h, is given by

$$\hat{X}_h(1) = \phi_0 + \sum_{i=1}^{p} \phi_i X_{h-i+1} \tag{2.9}$$

with associated forecast error $e_h(1) = X_{h+1} - \hat{X}_h(1) = Z_{h+1}$. In general, the ℓ-day-ahead forecast is given by

$$\hat{X}_h(\ell) = \phi_0 + \sum_{i=1}^{p} \phi_i \hat{X}_h(\ell - i). \tag{2.10}$$

It can be shown that $\hat{X}_h(\ell)$ converges to $\mathrm{E}(X_t)$ as ℓ tends to infinity; this is the so-called mean reversion of finance litterature (Tsay 2010, p. 56).

2.2.3 Conditional Heteroscedastic Models

We now move on to models for the variance of a time series. We look at models in which the (conditional) volatility is not fixed, but rather evolves over time. Regarding the volatility of financial asset returns, Tsay (2010, p. 111) has this to say:

> Although volatility is not directly observable, it has some characteristics that are commonly seen in asset returns. First, there exist volatility clusters (i.e. volatility may be high for certain time periods and low for other periods). Second, volatility evolves over time in a continuous manner—that is, volatility jumps are rare. Third, volatility does not diverge to infinity—that is, volatility varies within some fixed range. Statistically speaking, this means that volatility is often stationary. Fourth, volatility seems to react differently to a big price increase or a big price drop, referred to as the *leverage* effect.

These are good reasons to why the conditional volatility should be included when modelling asset returns, especially when the purpose is to estimate financial risk. One of the first models of this type was the autoregressive conditional heteroscedasticity (ARCH) model, due to Engle (1982). The idea behind the ARCH model is that the shocks ϵ_t of an asset return are dependent but serially uncorrelated, and that the dependence can be fully described by a function of its lagged values (Tsay 2010, pp. 116-117). To be precise, we may define an ARCH(p) process as follows:

Definition. *Let $\{Z_t\}$, be a strict white noise process with mean zero and variance one. The mean adjusted process $\{X_t - \mu\} = \{\epsilon_t\}$, is an ARCH(p) process if it is strictly stationary and if, for all t and for a strictly positive-valued process*

$\{\sigma_t\}$, it satisfies the equations

$$\begin{aligned} \epsilon_t &= \sigma_t Z_t && (2.11) \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2, && (2.12) \end{aligned}$$

where $\alpha_0 > 0$ and $\alpha_i \geq 0$, $i = 1, \ldots, p$.
(McNeil et al. 2005, p. 139)

From this model, it is seen that if any of $|\epsilon_{t-1}|, \ldots, |\epsilon_{t-p}|$ are large, then ϵ_t will be drawn from a distribution with large variance and is likely to also be large; this is how volatility clusters are generated (McNeil et al. 2005, pp. 139-140).

Though the ARCH model is simple, it is often the case that too many parameters need to be estimated in order to describe the volatility process of a returns series (Tsay 2010, p. 131). For this reason and others, several extensions and generalizations of the ARCH model have been made. There are several generalizations of the ARCH model. One such model is the aptly named Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model attributed to Bollerslev (1986). This model may be defined as follows:
Definition. *Let $\{Z_t\}$ be a strict white noise process with mean zero and variance one. The mean adjusted process $\{X_t - \mu\} = \{\epsilon_t\}$ is a GARCH(p, q) process if it is strictly stationary and if, for all t and for a strictly positive-valued process $\{\sigma_t\}$, it satisfies the equations*

$$\begin{aligned} \epsilon_t &= \sigma_t Z_t && (2.13) \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, && (2.14) \end{aligned}$$

where $\alpha_0 > 0$ and $\alpha_i \geq 0$, $i = 1, \ldots, p$, and $\beta_j \geq 0$, $j = 1, \ldots, q$.
(McNeil et al. 2005, p. 145)

The GARCH process is a generalized form of the ARCH process in the sense that the variance σ_t^2 may depend on earlier values of the variance, as well as the squares of earlier values of the process. We focus on the simplest GARCH(1,1) model, which by equation 2.14 has a conditional variance of the form

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_1^2 + \beta_1 \sigma_{t-1}^2, \tag{2.15}$$

where $\alpha_1 + \beta_1 < 1$; this condition guarantees that the mean adjusted series $\{\epsilon_t\}$ is strictly stationary (McNeil & Frey 2000, p. 277). This type of model can explain the persistence of high volatility and volatility clusters through the fact that $|\epsilon_t|$ can be large if either $|\epsilon_{t-1}|$ is large or σ_{t-1} is large (McNeil et al. 2005, p. 145). A high value of β_1 relative to α_1 means that volatility is persitent, and vice versa, a high value of α_1 relative to β_1 means that volatility is quick to react to market movements and appears 'spiky' (Dowd 2005, p. 132).

We can obtain forecasts of volatility from the GARCH(1,1) model with ease. Assuming the forecast origin is h, the one-step-ahead forecast is given by

$$\sigma_h^2(1) = \sigma_{h+1}^2 = \alpha_0 + \alpha_1 \epsilon_h^2 + \beta_1 \sigma_h^2 \tag{2.16}$$

where ϵ_h and σ_h are known at time index h (Tsay 2010, p. 133). By some clever substitutions, we can obtain the $\ell > 1$ day forecast, given by

$$\sigma_h^2(\ell) = \alpha_0 + (\alpha_1 + \beta_1)\sigma_h^2(\ell - 1) \tag{2.17}$$

It can furthermore be shown that if $(\alpha_1 + \beta_1) < 1$, then the ℓ-day-ahead forecast approaches the unconditional variance of ϵ_t as ℓ approaches infinity (Tsay 2010, p. 133).

We will use the forecasting methods of the AR(1) and GARCH(1,1) models later, when we employ conditional models to predict Value at Risk.

2.3 Extreme Value Theory

Extreme value theory (EVT) is a field of statistics concerned with the tail behavior of random variables, making it a suitable candidate for purposes of estimating tail-related risk measures such as Value at Risk. Much of the development of EVT has taken place during the early and mid parts of the 20th century, with significant contributions from Fréchet, Fisher and Tippett, and Gnedenko (Kotz & Nadarajah 2000, p. 2). Extreme value theory has applications in a wide array of fields, for which the common denominator is the question of how to deal with rare, high-impact events. To name a few examples, applications can be found in astronomy, analysis of natural disasters, insurance, and finance (Embrechts et al. 1997; Kotz & Nadarajah 2000). In this essay, we focus on the use of EVT to calculate value at risk for financial returns, and thus choose to cover only the theory which is relevant to this task. For a more comprehensive introduction to EVT, the reader is referred to de Haan & Ferreira (2006), and for applications of EVT to finance, Embrechts et al. (1997) is an oft-cited choice.

Two main approaches to the analysis of extremes can be found in the EVT litterature today. The first, based on the Fisher-Tippett-Gnedenko Theorem, concerns the limit distribution of centred and normalized sample maxima. The second approach, and the one used in this essay, considers instead the excesses of random variables over a high threshold. We thus begin by introducing the Generalized Pareto Distribution, with which this phenomena can be modelled:

The generalized Pareto distribution (Embrechts et al. 1997, p. 162) is a distribution function (d.f.) of the form

$$G_{\xi;\nu,\beta}(x) = \begin{cases} 1 - (1 + \xi \frac{x-\nu}{\beta})^{-1/\xi} & \text{if } \xi \neq 0, \\ 1 - \exp(-\frac{x-\nu}{\beta}) & \text{if } \xi = 0, \end{cases} \tag{2.18}$$

where $\beta > 0$ and $x \geq \nu$ if $\xi \geq 0$, and $\nu \leq x \leq \nu - \frac{\beta}{\xi}$ if $\xi < 0$. β is called a scale parameter, and governs the statistical dispersion of the distribution. ξ is called the shape parameter, and can be said to measure the heavy-tailedness of the distribution (McNeil & Saladin 1997). Specifically, the case $\xi < 0$ corresponds to distributions with a finite right endpoint, such as the uniform distribution. The case $\xi = 0$ corresponds to medium-tailed distributions, which include the normal and log-normal distributions. Lastly, the case $\xi > 0$ holds for heavy-tailed distributions such as the t-distribution, and is of high interest to us due to the apparent fat-tailedness of financial data.

In the following, we are mostly concerned with the case $G_{\xi;0,\beta}(y)$, where $y = x - u$, and we will denote this by $G_{\xi;\beta}(y)$. We also introduce the excess distribution function

$$F_u(y) = P(X - u \leq y | X > u) = \frac{F(y+u) - F(u)}{1 - F(u)}, y \geq 0 \tag{2.19}$$

Denoting the right tail of a given d.f. F as $\overline{F} = 1 - F$, we can also write this relationship as

$$\overline{F}_u(y) = P(X - u > y | X > u) = \frac{\overline{F}(y+u)}{\overline{F}(u)} \tag{2.20}$$

Pickands (1975) and Balkema & de Haan (1974) showed that, given a large threshold u, $F_u(y) \approx G_{\xi;\beta}(y)$ for a large class of distribution functions F, including, for example, the Pareto distribution, the uniform distribution, and the normal and log-normal distribution. It is this result we intend to use to model VaR.

2.3.1 An EVT-estimator for a quantile of the distribution

We now introduce what is known as the Peak-Over-Threshold (POT) model, which deals with observations above a certain threshold. We thus suppose that we have random variables $X_1, X_2, \ldots, X_n$ that are independent and identically distributed (i.i.d.), belonging to the class of distribution functions for which the result above holds. We choose a threshold u and define N_u to be the number of exceedances of u by $X_1, \ldots, X_n$. We denote these exceedances $Y_1, \ldots, Y_{N_u}$; that is, $Y_i = X_i - u$ for $i = 1, 2, \ldots, N_u$. We then wish to estimate parameters ξ and $\beta = \beta(u)$ from the data of excesses over u. According to Embrechts et al. (1997, p. 354) a good estimator for $\overline{F}(u)$ is the right tail of the empirical distribution function:

$$\widehat{\overline{F}(u)} = \overline{F}_n(u) = \frac{1}{n}\sum_{i=1}^{n} 1\{X_i > u\} = \frac{N_u}{n}, \tag{2.21}$$

where n is the size of the sample, and $1\{X \in A\}$ is an indicator variable, taking the value 1 if $X \in A$ and 0 otherwise, for an event A. Since $\overline{F}_u(y) \approx \overline{G}_{\xi,\beta(u)}(y)$, we can employ an estimator

$$\widehat{\overline{F}_u(y)} = \overline{G}_{\hat{\xi},\hat{\beta}}(y), \tag{2.22}$$

for the excess distribution function, where $\hat{\xi} = \hat{\xi}_{N_u}$ and $\hat{\beta} = \hat{\beta}_{N_u}$ can be obtained e.g. by maximum-likelihood estimation. Rearranging (2.20), we see that

$$\overline{F}(u+y) = \overline{F}(u)\overline{F}_u(y) \tag{2.23}$$

Putting these estimators together, we have a tail estimator for $y > 0$ of the form

$$\widehat{\overline{F}(u+y)} = \frac{N_u}{n}(1 + \hat{\xi}\frac{y}{\hat{\beta}})^{-1/\hat{\xi}} \tag{2.24}$$

However, of greater interest to us is the inverse of the above formula, for Value at Risk is essentially a quantile of the distribution of returns. For a chosen high probability q, simple calculation shows that an estimator for the q-quantile is

$$\hat{x}_q = u + \frac{\hat{\beta}}{\hat{\xi}}((\frac{n}{N_u}(1-q))^{-\hat{\xi}} - 1). \tag{2.25}$$

The theory we have covered applies equally well to the left tail of a distribution; in fact, we may go from modelling the left tail of the distribution of a random variable to modelling the right tail of the negative of that variable without any distortion of the results. For our purposes, we will simply negate the returns from our financial series and model VaR as an upper quantile of the distribution of losses. By doing so, the Peak-Over-Threshold Value at Risk may be estimated by (2.25). However, there are two main difficulties in applying this theory to financial data. First, the choice of threhold is critical for the estimation of the model, but this choice is not an easy one to make. Second, since financial returns are known not to be independent over time, we must either adapt the theory to fit the data, or adapt the data to fit the theory. We choose the latter approach, and follow McNeil & Frey (2000) in applying the POT model to the standardized residuals of an AR(1)-GARCH(1,1) model, as explained in section 3.3.2.

2.4 Historical Simulation

Historical simulation (HS) is one of the most intuitive and popular approaches for estimating VaR (Jorion 2001, p. 223). HS is a non-parametric approach, meaning that it can provide an estimate of VaR without requiring any specific assumptions about the theoretical distribution of the asset returns. The historical simulation approach to calculating VaR assumes that "all possible future variations have been experienced in the past, and that the historically simulated distribution is identical to the returns distribution over the forward looking risk horizon" (Alexander 2008, p. 42). The advantages of HS is that it is easy to implement, and that it, in the context of holding a portfolio consisting of many assets, "reduces the risk-measure estimation problem to a one-dimensional problem; no statistical estimation of the multivariate distribution of [risk factor changes] is necessary, and no assumptions about the dependence structure of risk-factor changes are made" (McNeil et al. 2005, p. 51). In essence, the computation of a covariance matrix (not discussed in this essay) is greatly simplified (Jorion 2001, p. 223).

Historical simulation is actually an umbrella term that covers several ways of calculating VaR. Since HS is only used as a comparison measure to the main subjects of this essay (conditional models and conditional EVT), we will only cover the simplest HS method here. For a more comprehensive coverage of the various methods that fall under the HS approach, see chapter 4 of Dowd (2005).

The method used in this essay is called *Basic Historical Simulation*, and can be implemented as follows: given a confidence level $q = 1 - \alpha$ and losses (i.e. negative returns) $x_{t_1}, \ldots, x_{t_n}$ for the last n days, order these observations in increasing order as $x_{(1)} \leq \ldots \leq x_{(n)}$. The VaR is then taken as the observation

x_q to the right of which $100\alpha\%$ of the observations lie (Dowd 2005, p. 84). For example, if $n = 1000$ and $q = 0.95$, then the VaR according to Basic HS would be the 51st largest loss. It is easy to imagine that extensions to this method may involve interpolation between the order statistics, but other methods also involve bootstrapping and other relatively advanced statistical techniques.

Lastly, it should be mentioned that the Basel Committee recommends using a sample period of daily data between 3 and 5 years when using historical simulation models to assess the regulatory market risk capital requirement (Alexander 2008, p. 145).

3 Methodology

3.1 Data

We apply the theory to three asset classes with three data sets in each. The data was collected from Thomson Reuters Datastream (Datastream), Wikiposit, and the U.S. Energy Information Administration (E.I.A.). The data sets from Wikiposit and the E.I.A. are freely available on the internet, while the data from Datastream can only be accessed with subscription. These are the data sets:

Asset	Acronym	Source	Currency	Start date	End date
Stock indices					
Nikkei 225	N225	Datastream	JPY	1980-01-01	2010-12-31
S&P 500 Composite	S&P500	Datastream	USD	1980-01-01	2010-12-31
DAX	DAX	Datastream	Euro	1980-01-01	2010-12-31
Stocks					
Ericsson B	ERICB	Datastream	SEK	1982-01-04	2010-12-31
Electrolux B	ELUXB	Datastream	SEK	1982-01-04	2010-12-31
Holmen B	HOLMB	Datastream	SEK	1982-01-04	2010-12-31
Commodities					
Crude Oil future	WTI	E.I.A.	USD	1983-04-04	2010-12-31
Gold spot price	GOLD	Wikiposit	USD	1980-01-01	2010-12-31
Wheat future	WHT	Wikiposit	USD	1980-01-01	2010-12-31

Table 1: List of data sets tested

Description of the data sets:

- The Nikkei 225 is a price-weighted average of stocks on the Tokyo Stock Exchange.
- The S&P 500 is "a capitalization-weighted index of 500 stocks intended to be a representative sample of leading companies in leading industries within the U.S. economy" (U.S. Securities and Exchange Commission n.d.).
- DAX, which stands for *Deutscher Aktien IndeX*, is a large cap index of 30 major companies trading on the Frankfurt Stock Exchange.
- Ericsson B, Electrolux B, and Holmen B are blue chip stocks for three major Swedish companies.

- The WTI consists of prices for futures contracts for West Texas Intermediate crude oil, in dollars per barrel. The delivery point is Cushing, Oklahoma, and the contracts are in units of 1000 barrels. The specific contract is the Cushing Crude Oil Future Contract 1. WTI is one of the major benchmarks for crude oil prices.
- GOLD is the London Afternoon Fixing for gold spot prices, in dollars per ounce.
- WHT is for the Kansas City Wheat Futures, specifically KC Wheat, 4th Contract.

3.2 Implementation

Code for the backtests and simulations of this essay were written in R; an open-source programming language and statistical software environment whose capabilities are extended through user-created *packages*, often geared toward a particular field or implementation. We used the package *timeSeries* (Wuertz & Chalabi 2010) for initial processing of the data; the package *fGarch* (Wuertz et al. 2009) for the conditional models; and the package *POT* (Ribatet 2011) to implement the EVT models. Plots were created with the package (and accompanying book) *ggplot2* (Wickham 2009). We happily share the code we've written to those who ask; just send an email to Benjamin about this. Altogether, the code is a few hundred lines, so we choose not to include it in an appendix.

3.3 Estimating VaR

Using the theory about stochastic processes and EVT covered in section 2, we will now explain how these results can be used to estimate Value at Risk. Our presentation follows that of McNeil & Frey (2000). Note that the method of estimating VaR using historical simulation was covered in section 2.4.

3.3.1 Estimating VaR using an AR(1)-GARCH(1,1) model

Our goal is to produce estimates of VaR that take into account the stochastic nature of volatility; in particular the volatility clustering that is often seen in asset return data. To this end, we assume that the (negative) log returns

$(X_t, t \in \mathbb{Z})$ of the assets analyzed form a strictly stationary process as

$$X_t = \mu_t + \sigma_t Z_t, \tag{3.1}$$

where

$$\begin{aligned} \mu_t &= \phi_0 + \phi_1 X_{t-1} && \text{(AR(1))} \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, && \text{(GARCH(1,1))} \end{aligned} \tag{3.2}$$

and $\epsilon_t = X_t - \mu_t$. As explained in section 2.2.3, the mean-adjusted series ϵ_t is strictly stationary if $\alpha_1 + \beta_1 < 1$, and this also implies that the distribution of the returns has a finite variance.

The innovations Z_t are taken to be i.i.d. with mean zero and unit variance, and so far, we make no assumptions about the marginal distribution function $F_Z(z)$. We let $\mathcal{K}_t$ represent the known information up to and including time t, and let

$$\begin{aligned} F_{X_{t+1}|\mathcal{K}_t}(x) &= P(\mu_{t+1} + \sigma_{t+1} Z_{t+1} \leq x|\mathcal{K}_t) \\ &= F_Z(\frac{x - \mu_{t+1}}{\sigma_{t+1}}) \end{aligned} \tag{3.3}$$

be the predictive distribution for the next day. Since we are only interested in the 1-day VaR, we limit ourselves to this case, although the concept can obviously be generalized. Conditional on $\mathcal{K}_t$, we can at day t get a quantile of the predictive distribution of the return for day $t+1$ as

$$\begin{aligned} x_q^{(t)} &= \inf\{x : F_{X_{t+1}|\mathcal{K}_t}(x) \geq q\} \\ &= \mu_{t+1} + \sigma_{t+1} z_q \end{aligned} \tag{3.4}$$

by (3.3). Since we assumed the innovations Z_t are i.i.d., the qth quantile z_q is independent of t. To estimate VaR, we must first choose the distribution of the innovations, so that the quantile z_q can be obtained. Then, we can proceed to estimate μ_{t+1} and σ_{t+1} conditional on the chosen distribution. We test two distributions for the innovations: a standard normal distribution, and a student-t distribution with $\nu = 4$ degrees of freedom, scaled to have variance 1. This choice of ν means that only the first three moments (expected value, variance, skewness) of the distribution of the innovations are defined, and should make for a good comparison to the normal distribution, for which all moments are finite. For the standard normal distribution, a quantile of the innovation distribution may be calculated as $z_q = \Phi^{-1}(q)$, where Φ is the d.f. of the standard normal distribution. The quantile for the t-distribution may be calculated as $z_q = \sqrt{(\nu - 2)/\nu} F_T^{-1}(q)$, where F_T is the distribution function for a t-distributed random variable with ν degrees of freedom. Quantiles for the standard normal distribution and the t-distribution may be found in common statistical tables, or, for more precision, be calculated by statistical software.

Assuming we have a data set of n observations ordered by time, we may through maximum likelihood estimation obtain estimates $\hat{\theta} = (\hat{\phi}_0, \hat{\phi}_1, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1)$ for the parameters of (3.2), and thus for $\hat{\mu}_i$ and $\hat{\sigma}_i, i = t-n+1, \ldots, t$ of equation (3.1). An estimate given at time t for VaR one day forward, with innovations following a standard normal distribution is then

$$\widehat{VaR}_q = \hat{x}_q^{(t+1)} = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\Phi^{-1}(q). \tag{3.5}$$

Similarly, the VaR for a return process where the innovations follow a student-t distribution with unit variance is given by

$$\widehat{VaR}_q = \hat{x}_q^{(t+1)} = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\sqrt{(\nu-2)/\nu}F_T^{-1}(q). \tag{3.6}$$

Given that the estimates $\hat{\theta}$ are the same for both estimators, the student-t estimate of the Value at Risk will always be higher than the estimate by the normally distributed innovations, for finite degrees of freedom ν. However, aside from the issue of whether the returns should be modelled as an AR(1)-GARCH(1,1) process to begin with, the choice of distribution for the innovations may be questioned. What if the true distribution has fatter tails still? It is for this reason we now consider what Extreme Value Theory has to offer.

3.3.2 Estimating VaR using conditional EVT

We now wish to replace the distribution of the tails of the innovations by the POT model discussed in section 2.3. Using the estimates $\hat{\mu}_i, \hat{\sigma}_i$, where $i = t-n+1, \ldots, t$, calculated for the AR(1)-GARCH(1,1) model in section 2.2 we may calculate the (standardized) residuals by

$$(z_{t-n+1}, \ldots, z_t) = (\frac{x_{t-n+1} - \hat{\mu}_{t-n+1}}{\hat{\sigma}_{t-n+1}}, \ldots, \frac{x_t - \hat{\mu}_t}{\hat{\sigma}_t}). \tag{3.7}$$

Given the n calculated residuals of (3.7) and a threshold u, we may first use equation (2.25) to estimate a quantile for the distribution of the innovations as

$$\hat{z}_q = u + \frac{\hat{\beta}}{\hat{\xi}}((\frac{n}{N_u}(1-q))^{-\hat{\xi}} - 1). \tag{3.8}$$

With this estimate at hand, we may then obtain an EVT-based estimate of the VaR as

$$\widehat{VaR}_q = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\hat{z}_q. \tag{3.9}$$

The question of how to choose the threshold u in equation (3.8) remains. Here, we follow McNeil & Frey (2000) in choosing to fix the number N_u of exceedances of the threshold to a value k, and simply let the threshold be the $(k+1)$th largest residual. To choose k, we perform a Monte Carlo simulation to estimate the bias and mean square error (MSE) of the estimators $\hat{\xi}$ and $\hat{z}_q$. If we can find a value k for which the bias and MSE are stable and satisfactory, we can then fix the value k throughout the backtests we perform later on. This would greatly simplify the procedure.

3.4 Choice of threshold

Before we can begin the backtesting procedures, we must choose a threshold for the GPD model. As explained in section 3.3.2, we choose the threshold by fixing a number k of observations to lie in the upper tail of the distribution, and let the threshold be the $(k+1)$th order statistic. Following McNeil & Frey (2000, pp. 284-287), we perform a Monte Carlo simulation to measure the bias and mean square error (MSE) of $\hat{\xi}$ and $\hat{z}_q$, to see if we can choose k so that our estimates are stable. Treating k as fixed (so that we may omit its mention in the notation), and treating the estimators as random variables, we can per de Haan & Ferreira (2006, p. 148) define an estimator of the MSE as

$$\widehat{\text{MSE}}(\hat{z}_q) = \frac{1}{r}\sum_{i=1}^{n}(\hat{z}_q^{(i)} - z_q)^2. \tag{3.10}$$

$\hat{z}_q^{(i)}$ is here the quantile estimate for the ith sample. We may estimate the bias of $\hat{z}_q$ as

$$\widehat{\text{Bias}}(\hat{z}_q) = \bar{\hat{z}}_q - z_q, \tag{3.11}$$

where $\bar{\hat{z}}_q$ is the average of the estimates of z_q (for k fixed), since it is easy to see that

$$\begin{aligned}
\text{E}(\widehat{\text{Bias}}(\hat{z}_q)) &= \text{E}(\bar{\hat{z}}_q - z_q) \\
&= \text{E}(\frac{1}{r}\sum_{i=1}^{r}\hat{z}_q^{(i)}) - \text{E}(z_q) \\
&= \frac{1}{r}(\sum_{i=1}^{r}\text{E}(\hat{z}_q^{(i)})) - z_q \\
&= \frac{1}{r}r\text{E}(\hat{z}_q) - z_q \\
&= \text{E}(\hat{z}_q) - z_q \\
&= \text{Bias}(\hat{z}_q),
\end{aligned}$$

i.e. an unbiased estimator of the bias. We focus on $\hat{z}_q$ here, but the calculations for $\hat{\xi}$ are similar. As explained by Coles (2001, p. 28), we are better off trying to minimize the MSE, since "the MSE measures variation of the estimator around the true parameter value, low MSE implies that in any particular sample the estimate is likely to be close to the true parameter value".

The simulation is implemented as follows: we draw $r = 1000$ samples of $n = 1000$ observations from a t-distribution with $\nu = 4$ degrees of freedom, and calculate estimates of bias and MSE from these for each value of k, holding q fixed. We do this for $q = 0.95,\ 0.99,\ 0.995$; these are the confidence levels used for the backtests later on. In figure 1, we show the plots for two such simulations.

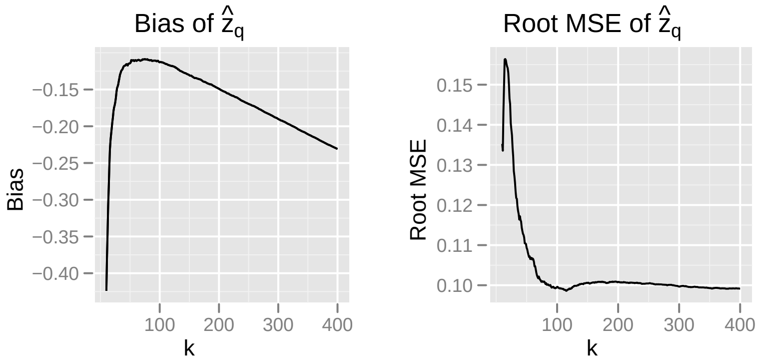

Figure 1: Monte Carlo estimates of bias (left) and root MSE (right) of $\hat{z}_q$ against different values of k, for the POT estimator of the 0.99 quantile of a t-distribution with four degrees of freedom.

Plots indicate that setting $k = 100$ provides relatively stable estimates, particularly for high quantiles. This is the same result found by McNeil & Frey (2000).

3.5 Backtesting

One way to test the efficacy of a VaR-model is to see how well it performs on historical data, and this can be done in a process known as backtesting. Backtesting can be defined as "the application of quantitative methods to determine whether the forecasts of a VaR forecasting model are consistent with the assumptions on which the model is based" (Dowd 2005, p. 321). The backtest is only a part of the model validation process, which may also include stress testing and independent review of model selection and implementation (Jorion

2001, p. 129). The Basel committee requires banks with permission to use internal models to calculate capital requirements to employ backtesting procedures for their risk measures (Jorion 2001, p. 129), and the specifications of these requirements have been subject to some revision after the financial crisis that started in 2007; see for example Basel Committee on Banking Supervision (2010).

A number of different backtesting methods exist, their suitability depending on the risk measure used, the structure of the portfolio, and many other factors. In this essay, we employ two backtesting procedures: the test of Christoffersen (1998), and, mostly for comparison with McNeil & Frey (2000), an exact binomial test. The Christoffersen backtest actually consists of three test statistics, so that three separate tests can be made. We describe each in turn. Last, we describe the exact binomial test.

3.5.1 Unconditional coverage tests

The basic idea for the unconditional coverage test is to treat each realized loss (return) as a Bernoulli-distributed random variable that exceeds a given level—the VaR—with probability $\alpha = 1 - q$; such an exceedance is called a VaR break. The number of VaR breaks for the estimation period (the period of the backtest) is then counted, and this number should thus be a random variable with a binomial distribution. Formally, we define an indicator function for a VaR break at time t as

$$I_t = 1\{X_t > x_q^{(t-1)}\} \sim \text{Bernoulli}(\alpha). \tag{3.12}$$

The null hypothesis is then that $\{I_t\} \sim$ i.i.d. Bernoulli(α). The density function is given by $f(k;\alpha) = \alpha^k(1-\alpha)^{1-k}$, $k \in \{0,1\}$, and if we let π denote the proportion of VaR-breaks in the sample (of length m), the Likelihood function of a Bernoulli(π) hit sequence may be written

$$\text{L}(\pi) = \prod_{t=1}^{m}(1-\pi)^{1-I_t}\pi^{I_t} = (1-\pi)^{m_0}\pi^{m_1}, \tag{3.13}$$

where m_0 is the number of 0's (non-VaR-breaks) and m_1 is the number of 1's (VaR breaks) in the sample (Christoffersen 2003, pp. 184-185). π is then easily estimated by $\hat{\pi} = m_1/m$; this is obviously the maximum likelihood estimate. Plugging this value into the Likelihood function then gives the optimized likelihood as

$$L(\hat{\pi}) = \prod_{t=1}^{m} (1-\hat{\pi})^{1-I_t} \hat{\pi}^{I_t} = (1-\hat{\pi})^{m_0} \hat{\pi}^{m_1}. \tag{3.14}$$

Under the null hypothesis, we have $\pi = \alpha$ (i.e. that the model predicts the correct proportion of VaR breaks), so that

$$L(\alpha) = \prod_{t=1}^{m} (1-\alpha)^{1-I_t} \alpha^{I_t} = (1-\alpha)^{m_0} \alpha^{m_1}. \tag{3.15}$$

The unconditional coverage hypothesis may then be tested by a likelihood ratio test of the form

$$LR_{uc} = -2 \ln [L(\alpha)/L(\hat{\pi})], \tag{3.16}$$

which is asymptotically distributed as a χ^2 variable with one degree of freedom (Christoffersen 2003, p. 185). This specific test is also known as the Kupiec backtest (Jorion 2001, p. 134), and the LR_{uc} statistic is one of two used for the conditional coverage test invented by Christoffersen (1998), covered two sections ahead.

3.5.2 Independence test

The unconditional coverage has the disadvantage that it does not account for clustering of VaR breaks; a phenomenon that may happen if the VaR-model does not respond sufficiently after a VaR break has occured. Christoffersen (1998) and Christoffersen (2003) describe a way in which the independence of the VaR breaks may be tested. For $i, j \in \{0, 1\}$, let $\pi_{ij} = P(I_t = j | I_{t-1} = i)$, i.e. the probability that state j occurs given that state i occured the day before. For example, π_{11} is the probability of a VaR break today given that there was a VaR break yesterday. This is a first-order Markov property: tomorrow's outcome only depends on the outcome today (Christoffersen 2003, p. 187). Thus, $\{I_t\}$ can be viewed as a binary first-order Markov chain with associated probability transition matrix $\mathbf{\Pi}_1$. Avoiding writing out the matrices, we skip directly to the punchline: this process has a likelihood function of the form

$$L(\mathbf{\Pi}_1) = (1-\pi_{01})^{m_{00}} \pi_{01}^{m_{01}} (1-\pi_{11})^{m_{10}} \pi_{11}^{m_{11}}, \tag{3.17}$$

where m_{ij} denotes the number of times in the sample a j was preceded by an i. The maximum likelihood estimates are given by

$$\hat{\pi}_{ij} = \frac{m_{ij}}{m_{ij} + m_{ii}}, \tag{3.18}$$

and plugging these values into (3.17) gives us an optimized likelihood $\mathrm{L}(\hat{\mathbf{\Pi}}_1)$. If the VaR breaks are independent over time, then we have $\pi_{01} = \pi_{11} = \pi$ and $\pi_{00} = \pi_{10} = 1 - \pi$. Denoting the associated probability transition matrix by $\mathbf{\Pi}$, this gives us a likelihood function

$$\mathrm{L}(\mathbf{\Pi}) = (1 - \pi)^{(m_{00}+m_{10})}\pi^{(m_{01}+m_{11})}. \tag{3.19}$$

Replacing π by the sample estimate

$$\hat{\pi} = \frac{m_{01} + m_{11}}{m_{00} + m_{10} + m_{01} + m_{11}} \tag{3.20}$$

we obtain an optimized likelihood $\mathrm{L}(\hat{\mathbf{\Pi}})$. Similar to the unconditional coverage test, we may form a likelihood ratio test of the form

$$LR_{ind} = -2 \ln [\mathrm{L}(\hat{\Pi})/\mathrm{L}(\hat{\Pi}_1)], \tag{3.21}$$

which is asymptotically distributed as a χ^2 variable with one degree of freedom (Christoffersen 1998, p. 846).

3.5.3 Conditional coverage test

Finally, it was the idea of Christoffersen (1998) to combine the two prior test statistics into one, to simultaneously test for accuracy in the prediction of the proportion of VaR breaks, and the independence of these VaR breaks. The test statistic is given by $\mathrm{LR}_{cc} = \mathrm{LR}_{uc} + \mathrm{LR}_{ind}$, and is asymptotically distributed as a χ^2 variable with two degrees of freedom (Christoffersen 1998, p. 846). Dowd (2005, p. 329) notes that

> The Christoffersen approach enables us to test both coverage and independence hypotheses at the same time. Moreover, if the model fails a test of both hypotheses combined, his approach enables us to test each hypothesis separately, and so establish where the model failure arises (e.g., does the model fail because of incorrect coverage, or does it fail because of lack of independence?).

For this reason, we provide the obtained values for all three statistics with associated p-values in section 5.

3.5.4 Exact binomial test

We also include an exact binomial test in our backtests, mostly to compare our results to those of (McNeil & Frey 2000). Just as for the unconditional coverage test, each VaR break is treated as a Bernoulli-distributed random variable with probability $\alpha = 1 - q$ of a VaR break. The number of VaR breaks for the estimation period (the period of the backtest) is then counted, and this number should thus be a random variable with a binomial distribution (since a sum of Bernoulli-distributed random variables is a binomially distributed random variable). Formally, if we assume that the return process as modelled by equation (3.1) holds true, then we may define an indicator function for a VaR break at time t as

$$I_t = 1\{X_{t+1} > x_q^{(t)}\} = 1\{Z_{t+1} > z_q\} \sim \text{Bernoulli}(\alpha). \tag{3.22}$$

Since Z_i and Z_j are assumed to be independent for $i \neq j$, we see that the number of VaR breaks B_m is distributed as

$$B_m = \sum_{t \in T} I_t \sim \text{Binomial}(m, \alpha), \tag{3.23}$$

where T is the set of the m days corresponding to the estimation period. We first observe that $\mathrm{E}(B_m) = \alpha m$, or equivalently, $\mathrm{E}(B_m/m) = \alpha$, so large deviations from these values should indicate that our model is incorrect. In essence, we will use a two-sided test to see if the proportion of exceedances B_m/m is significantly different from the true probability α. Under the null hypothesis that the model being tested is accurate, this difference should not be statistically significant.

4 Exploratory data analysis

Before proceeding with the backtests, it may be a good idea to inspect the data sets used, to see if they display some of the qualities that our models rely on them to do. Specifically, it is worth investigating the presence of fat-tailedness and autocorrelation in the data, and this can be done graphically through e.g. Q-Q plots and correlograms. Of course, if the data sets and time periods used for the investigation are the same ones to be used for backtesting, it is important not to let the conclusions drawn from the investigation affect model and parameter choices. In truth, the models used in this essay were chosen *a priori*, so the exploratory data analysis will be more of a check on realism than a source of input for model selection.

We limit ourselves to only analyzing the data set WTI here.

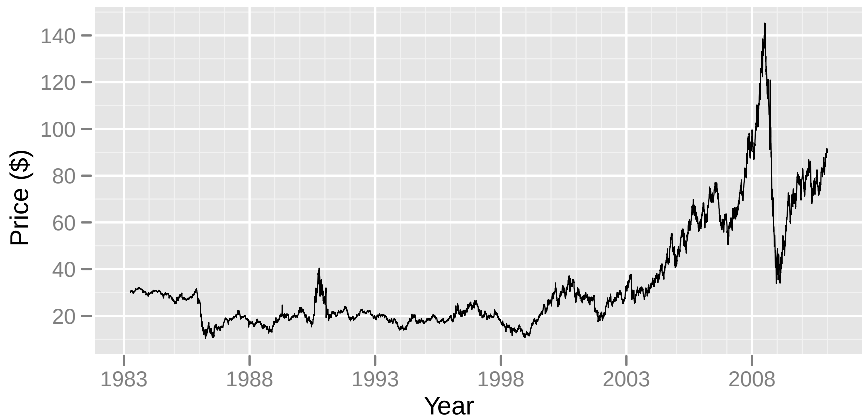

Figure 2: Price series for Western Texas Intermediate crude oil.

In figure 2, we show the price evolution of the WTI. We note that there seem to be some trends in the data, and the price drop during 2008 appears quite impressive. How does the return series look? Since we are interested in measuring risk, we will look at losses as positive amounts. This simply means that we calculate the log returns as usual, and then multiply these by -1.

In figure 3 below, we show the loss series for the WTI. We have marked the date 17 January 1991 in the plot; the official start date of Operation Desert Storm, and also the day when stockpiles of crude oil were released from the U.S. Strategic Petroleum Reserve. Although not immediately obvious in figure 2, crude oil prices fell by \$10.56 a barrel on this day. We also note that the

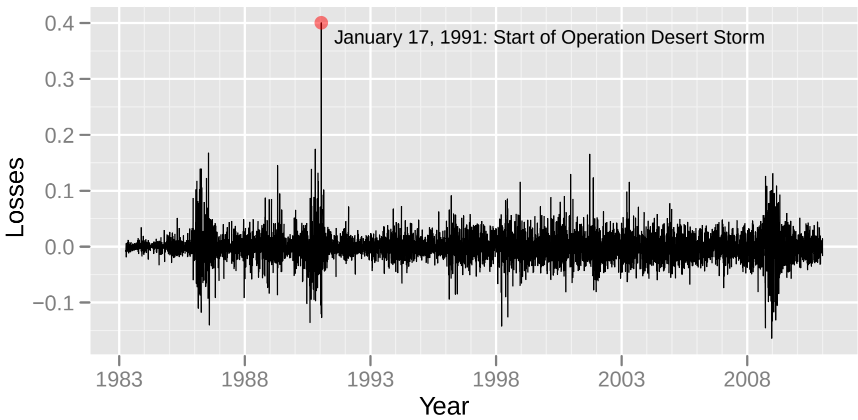

Figure 3: Log losses (negative log returns) for the WTI series.

returns seem more volatile during certain periods compared to 'the average'. We can investigate this further by plotting the conditional volatilty estimated from the AR(1)-GARCH(1,1) model, as done in figure 4. Comparing figure 4 to

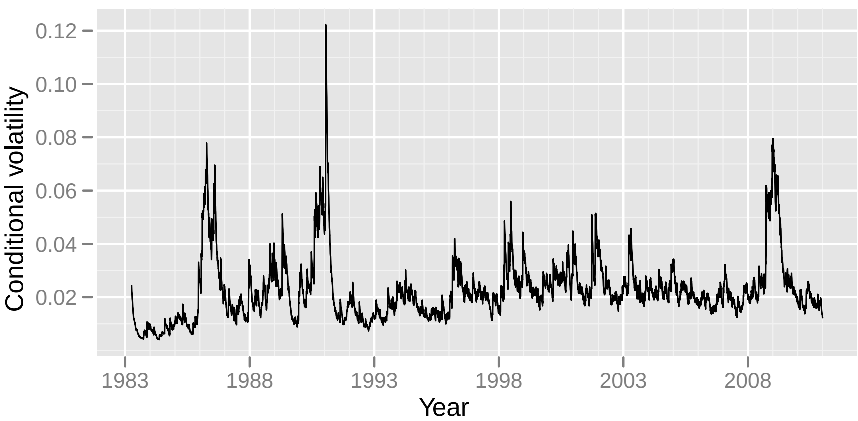

Figure 4: Conditional volatilty for the WTI series.

figure 3, we see that conditional volatilty indeed varies over time, being much higher than usual during certain periods.

Another interesting comparison is to the normal distribution. As pointed out in sections 1.1 and 1.4, financial data often has thicker tails than what would be the case under the normal distribution, and we should therefore be able to see this by comparing the quantiles of the normal distribution to those of our data sets. This can be done using a Q-Q plot: if the empirical distribution is the same as the theoretical distribution, then the observations should lie on a straight line. If the empirical distribution has thicker tails than the theoretical distribution, then the observations should lie below the line in the left tail, and above the line in the right tail. In figure 5, we compare both the observed losses and the residuals from the fitted AR(1)-GARCH(1,1) model to the standard normal distribution. To make the comparision easier, we standardize the losses by subtracting the sample mean and dividing by the sample standard deviation.

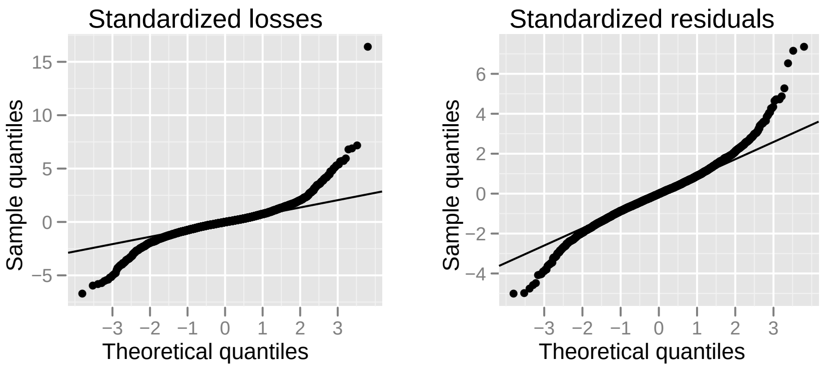

Figure 5: On the left: a Q-Q plot of standardized log losses versus a standard normal distribution. On the right: a Q-Q plot of standardized residuals from an AR(1)-GARCH(1,1) model versus a standard normal distribution.

As is seen, the fat tails remain even after fitting the AR(1)-GARCH(1,1) model to the data; this justifies somewhat the use of Extreme Value Theory to model the tails of the distribution of the innovations.

Finally, we want to investigate the autocorrelation in the data. This can be done graphically with an autocorrelation plot, also known as a correlogram. In figure 6, we show the values of the autocorrelation function for a range of lags (in days). An observation beyond the 95% confidence band, shown as the

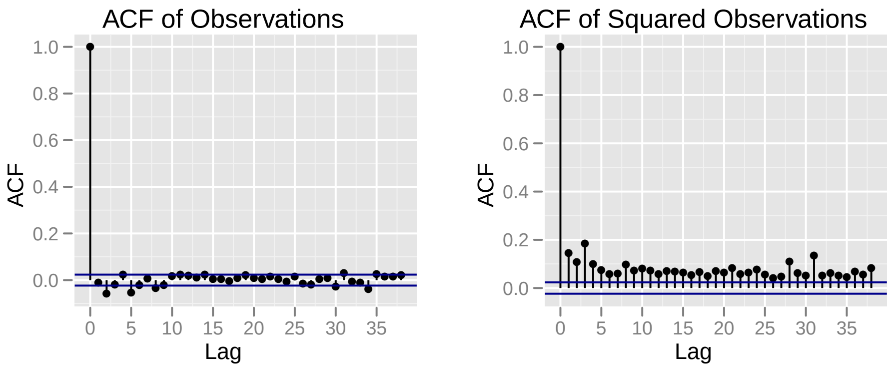

Figure 6: Correlogram for the observed losses (left) and their squares (right). Blue horizontal lines represent a 95% confidence band.

line in blue, can be interpreted as a sign of autocorrelation at the given lag. As is seen in figure 6, particularly the squared log losses show definite signs of autocorrelation at least 40 days back. By fitting an autoregressive model to the data, the autocorrelation should lessen. In figure 7, we see just that. This results speaks in favor of filtering the data with an autoregressive model before calculating VaR.

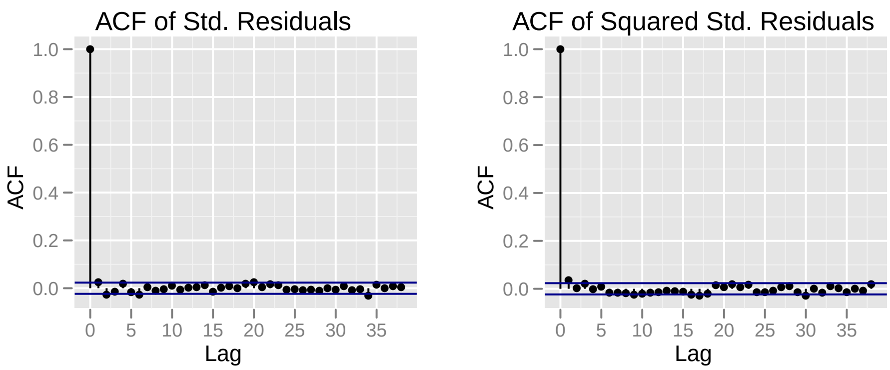

Figure 7: Correlogram for the standardized residuals (left) and their squares (right). Blue horizontal lines represent a 95% confidence band.

Thus, we now proceed to implement the different models and backtest them.

5 Results

5.1 Backtesting results

On the next few pages, we show the numerical results of the backtests performed, followed by a summary of what information can be extracted from these tables. The numbers used for the backtests are generated simply by stepping through the data sets day by day, using the previous 1000 days to estimate the parameters for the different models, and producing a VaR estimate for each model. This means that VaR is not estimated for the first 1000 days of each data set. Note that 'confidence level' refers to the confidence level of the VaR (i.e. q). We use a significance level of 0.05 for deciding whether a model is to be rejected or not for all the backtests, although we also provide p-values in the tables.

The results can be represented visually. Below, we show the VaR-estimates for two of the models, for a select period of time of the WTI data set.

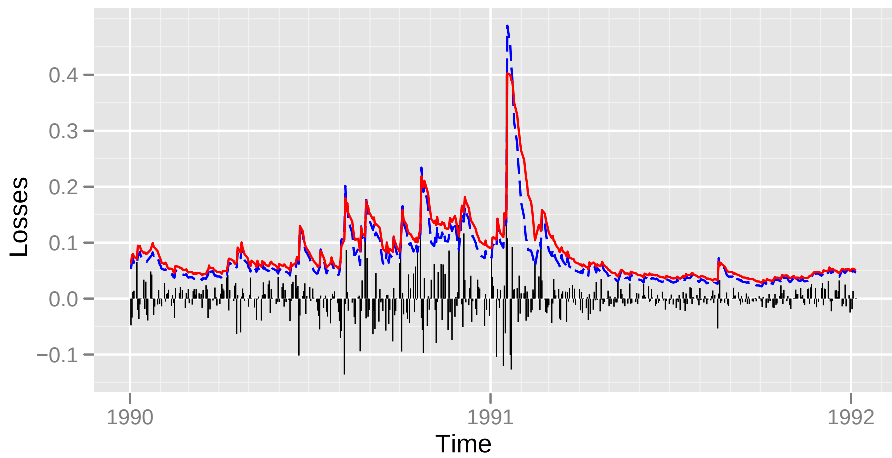

Figure 8: 99% confidence level VaR estmates for part of the WTI series. The blue dashed line shows the estimates of an AR(1)-GARCH(1,1) model with conditional normal innovations. The red solid line shows estimates of an AR(1)-GARCH(1,1) model with conditional t-distributed innovations, the tails of which have been modelled with EVT.

As can be seen in figure 8, the reaction to volatility spikes is less severe for the conditional EVT model than for the conditional normal model, but the increased VaR level also subsides slower. Overall, the conditional EVT model gives higher VaR estimates than the conditional normal model, which is to be expected. It is clear that both models respond quickly to increases in volatility; an unconditional model would be slower to react.

Dataset	N225	S&P500	DAX	ERICB	ELUXB	HOLMB	WTI	GOLD	WHT	Times
Length of test	7088	7086	7088	6564	6564	6564	6174	6776	6802	rejected
0.95 Quantile										
Unconditional EVT	63.7 (0.000)	49.8 (0.000)	56.1 (0.000)	48.8 (0.000)	38.1 (0.000)	38.7 (0.000)	32.7 (0.000)	29.9 (0.000)	25.6 (0.000)	9
Conditional t	55.9 (0.000)	26.9 (0.000)	71.7 (0.000)	4.3 (0.115)	5.1 (0.076)	11.3 (0.004)	29.3 (0.000)	11.3 (0.004)	3.2 (0.202)	6
Conditional EVT (t)	7.2 (0.027)	2.2 (0.341)	4.2 (0.125)	3.9 (0.144)	10.2 (0.006)	0.4 (0.835)	4.4 (0.113)	9.8 (0.008)	6.1 (0.048)	4
Conditional normal	13.5 (0.001)	1.2 (0.537)	6.4 (0.041)	7.8 (0.020)	7.3 (0.026)	6.2 (0.046)	5.7 (0.058)	11.8 (0.003)	22.5 (0.000)	7
Conditional EVT (n)	4.8 (0.090)	0.9 (0.627)	1.9 (0.396)	4.4 (0.111)	8.4 (0.015)	0.1 (0.933)	4.3 (0.119)	8.2 (0.017)	2.7 (0.254)	2
Historical	56.1 (0.000)	25.8 (0.000)	49.1 (0.000)	49.5 (0.000)	35.1 (0.000)	35.0 (0.000)	27.1 (0.000)	17.7 (0.000)	24.9 (0.000)	9
0.99 Quantile										
Unconditional EVT	64.7 (0.000)	117.8 (0.000)	52.5 (0.000)	19.5 (0.000)	33.3 (0.000)	38.6 (0.000)	31.1 (0.000)	73.4 (0.000)	37.1 (0.000)	9
Conditional t	1.4 (0.495)	2.8 (0.242)	1.5 (0.467)	1.3 (0.511)	1.6 (0.448)	3.7 (0.158)	6.9 (0.032)	6.2 (0.045)	1.3 (0.509)	2
Conditional EVT (t)	0.1 (0.948)	1.5 (0.466)	1.5 (0.466)	1.4 (0.494)	1.8 (0.413)	2.3 (0.313)	2.2 (0.327)	5.8 (0.054)	0.9 (0.626)	0
Conditional normal	48.2 (0.000)	58.6 (0.000)	33.2 (0.000)	34.7 (0.000)	11.2 (0.004)	21.2 (0.000)	18.3 (0.000)	14.1 (0.001)	9.7 (0.008)	9
Conditional EVT (n)	0.1 (0.948)	2.2 (0.340)	0.1 (0.947)	2.4 (0.294)	1.8 (0.413)	1.8 (0.407)	1.5 (0.465)	6.6 (0.037)	0.2 (0.925)	1
Historical	26.2 (0.000)	52.2 (0.000)	34.3 (0.000)	16.1 (0.000)	26.1 (0.000)	39.7 (0.000)	36.0 (0.000)	23.5 (0.000)	38.7 (0.000)	9
0.995 Quantile										
Unconditional EVT	77.1 (0.000)	181.4 (0.000)	69.9 (0.000)	15.4 (0.000)	37.0 (0.000)	28.1 (0.000)	23.6 (0.000)	106.1 (0.000)	25.0 (0.000)	9
Conditional t	3.3 (0.191)	3.8 (0.149)	5.0 (0.081)	0.5 (0.772)	0.3 (0.846)	5.7 (0.057)	4.1 (0.126)	3.7 (0.158)	1.4 (0.505)	0
Conditional EVT (t)	0.5 (0.787)	2.9 (0.235)	2.9 (0.235)	0.3 (0.846)	2.8 (0.245)	2.3 (0.319)	2.1 (0.344)	0.3 (0.842)	2.3 (0.320)	0
Conditional normal	50.7 (0.000)	71.1 (0.000)	43.1 (0.000)	52.4 (0.000)	26.5 (0.000)	42.6 (0.000)	24.0 (0.000)	31.1 (0.000)	18.5 (0.000)	9
Conditional EVT (n)	0.6 (0.728)	1.8 (0.404)	2.3 (0.321)	0.4 (0.820)	3.6 (0.169)	2.1 (0.351)	0.3 (0.858)	0.5 (0.773)	0.9 (0.644)	0
Historical	20.0 (0.000)	37.2 (0.000)	33.2 (0.000)	9.9 (0.007)	30.6 (0.000)	21.1 (0.000)	19.3 (0.000)	1.9 (0.385)	20.0 (0.000)	8

Table 2: Backtesting results from the conditional coverage test. Tabel entries show LR_{cc} statistics (χ^2 distributed with 2 degrees of freedom), and associated p-value in parentheses. Rightmost column shows number of times (out of 9) the model is rejected at significance level 0.05.

Dataset	N225	S&P500	DAX	ERICB	ELUXB	HOLMB	WTI	GOLD	WHT	Times
Length of test	7088	7086	7088	6564	6564	6564	6174	6776	6802	rejected
0.95 Quantile										
Unconditional EVT	24.7 (0.000)	26.1 (0.000)	24.1 (0.000)	13.1 (0.000)	15.5 (0.000)	1.6 (0.201)	6.9 (0.009)	8.1 (0.004)	13.8 (0.000)	8
Conditional t	54.8 (0.000)	24.1 (0.000)	71.6 (0.000)	1.2 (0.267)	1.0 (0.317)	7.5 (0.006)	24.1 (0.000)	4.8 (0.028)	0.1 (0.743)	6
Conditional EVT (t)	3.9 (0.049)	1.2 (0.266)	4.1 (0.044)	1.9 (0.165)	5.1 (0.023)	0.1 (0.743)	1.5 (0.215)	0.1 (0.772)	4.3 (0.038)	4
Conditional normal	12.9 (0.000)	0.9 (0.341)	5.0 (0.026)	7.5 (0.006)	3.9 (0.049)	5.7 (0.017)	2.9 (0.088)	4.4 (0.037)	21.5 (0.000)	7
Conditional EVT (n)	4.3 (0.038)	0.4 (0.529)	1.4 (0.244)	1.5 (0.222)	6.2 (0.013)	0.0 (0.919)	2.7 (0.100)	0.2 (0.663)	1.6 (0.207)	2
Historical	15.2 (0.000)	11.7 (0.001)	19.8 (0.000)	8.2 (0.004)	9.5 (0.002)	1.2 (0.267)	1.8 (0.175)	0.0 (0.859)	8.3 (0.004)	6
0.99 Quantile										
Unconditional EVT	49.0 (0.000)	98.8 (0.000)	41.2 (0.000)	10.2 (0.001)	20.2 (0.000)	5.3 (0.022)	8.3 (0.004)	62.5 (0.000)	21.1 (0.000)	9
Conditional t	1.4 (0.237)	2.0 (0.159)	0.1 (0.801)	0.0 (0.964)	1.2 (0.273)	3.1 (0.079)	5.8 (0.016)	0.7 (0.401)	0.0 (0.901)	1
Conditional EVT (t)	0.0 (0.987)	0.1 (0.712)	0.1 (0.712)	0.0 (0.866)	0.3 (0.593)	1.8 (0.175)	1.6 (0.200)	0.5 (0.476)	0.9 (0.340)	0
Conditional normal	46.3 (0.000)	55.8 (0.000)	32.9 (0.000)	30.6 (0.000)	9.5 (0.002)	21.1 (0.000)	17.1 (0.000)	12.7 (0.000)	9.6 (0.002)	9
Conditional EVT (n)	0.0 (0.987)	1.1 (0.286)	0.0 (0.894)	0.8 (0.370)	0.3 (0.593)	0.0 (0.964)	0.5 (0.460)	0.9 (0.334)	0.1 (0.810)	0
Historical	10.7 (0.001)	32.9 (0.000)	22.4 (0.000)	4.8 (0.029)	14.8 (0.000)	10.2 (0.001)	9.0 (0.003)	2.5 (0.117	9.0 (0.003)	8
0.995 Quantile										
Unconditional EVT	70.8 (0.000)	174.3 (0.000)	59.1 (0.000)	12.5 (0.000)	18.2 (0.000)	12.5 (0.000	12.5 (0.000)	105.3 (0.000)	19.8 (0.000)	9
Conditional t	0.9 (0.348)	1.3 (0.263)	2.2 (0.138)	0.1 (0.706)	0.0 (0.975)	2.6 (0.105)	1.2 (0.271)	1.1 (0.296)	1.1 (0.286)	0
Conditional EVT (t)	0.2 (0.679)	0.6 (0.445)	0.6 (0.445)	0.0 (0.885)	1.5 (0.225)	0.1 (0.748)	0.0 (0.969)	0.0 (0.984)	1.8 (0.185)	0
Conditional normal	50.0 (0.000)	70.7 (0.000)	43.1 (0.000)	50.3 (0.000)	24.6 (0.000)	41.6 (0.000)	23.8 (0.000)	26.7 (0.000)	17.4 (0.000)	9
Conditional EVT (n)	0.3 (0.557)	0.0 (0.925)	0.2 (0.678)	0.1 (0.748)	2.4 (0.124)	0.0 (0.885)	0.0 (0.969)	0.1 (0.718)	0.5 (0.501)	0
Historical	14.2 (0.000)	33.5 (0.000)	13.1 (0.000)	6.2 (0.013)	7.0 (0.008)	9.6 (0.002)	5.9 (0.015)	1.4 (0.235)	8.2 (0.004)	8

Table 3: Backtesting results from the unconditional coverage test. Tabel entries show LR_{uc} statistics (χ^2 distributed with 1 degree of freedom), and associated p-value in parentheses. Rightmost column shows number of times (out of 9) the model is rejected at significance level 0.05.

Dataset	N225	S&P500	DAX	ERICB	ELUXB	HOLMB	WTI	GOLD	WHT	Times
Length of test	7088	7086	7088	6564	6564	6564	6174	6776	6802	rejected
0.95 Quantile										
Unconditional EVT	39.1 (0.000)	23.7 (0.000)	32.0 (0.000)	35.8 (0.000)	22.6 (0.000)	37.1 (0.000)	25.8 (0.000)	21.8 (0.000)	11.8 (0.001)	9
Conditional t	1.1 (0.287)	2.8 (0.094)	0.0 (0.849)	3.1 (0.079)	4.1 (0.042)	3.8 (0.051)	5.2 (0.022)	6.4 (0.011)	3.1 (0.079)	3
Conditional EVT (t)	3.3 (0.068)	0.9 (0.338)	0.1 (0.757)	2.0 (0.162)	5.0 (0.025)	0.3 (0.615)	2.8 (0.093)	9.7 (0.002)	1.8 (0.183)	2
Conditional normal	0.6 (0.422)	0.3 (0.562)	1.4 (0.231)	0.3 (0.562)	3.4 (0.065)	0.5 (0.482)	2.8 (0.094)	7.5 (0.006)	0.9 (0.334)	1
Conditional EVT (n)	0.5 (0.477)	0.5 (0.463)	0.5 (0.483)	2.9 (0.089)	2.3 (0.133)	0.1 (0.719)	1.5 (0.214)	8.0 (0.005)	1.2 (0.283)	1
Historical	41.0 (0.000)	14.1 (0.000)	29.3 (0.000)	41.3 (0.000)	25.6 (0.000)	33.8 (0.000)	25.2 (0.000)	17.7 (0.000)	16.7 (0.000)	9
0.99 Quantile										
Unconditional EVT	15.7 (0.000)	19.0 (0.000)	11.3 (0.001)	9.3 (0.002)	13.1 (0.000)	33.3 (0.000)	22.8 (0.000)	10.9 (0.001)	16.0 (0.000)	9
Conditional t	0.0 (0.939)	0.9 (0.355)	1.5 (0.227)	1.3 (0.247)	0.4 (0.524)	0.6 (0.434)	1.1 (0.305)	5.5 (0.019)	1.3 (0.248)	1
Conditional EVT (t)	0.1 (0.745)	1.4 (0.238)	1.4 (0.238)	1.4 (0.240)	1.5 (0.224)	0.5 (0.488)	0.6 (0.443)	5.3 (0.021)	0.0 (0.872)	1
Conditional normal	1.9 (0.170)	2.8 (0.095)	0.3 (0.587)	4.1 (0.043)	1.7 (0.191)	0.0 (0.827)	1.2 (0.265)	1.3 (0.251)	0.1 (0.764)	1
Conditional EVT (n)	0.1 (0.745)	1.0 (0.313)	0.1 (0.764)	1.6 (0.200)	1.5 (0.224)	1.8 (0.180)	1.0 (0.320)	5.7 (0.017)	0.1 (0.753)	1
Historical	15.5 (0.000)	19.4 (0.000)	11.9 (0.001)	11.4 (0.001)	11.3 (0.001)	29.5 (0.000)	27.0 (0.000)	21.0 (0.000)	29.7 (0.000)	9
0.995 Quantile										
Unconditional EVT	6.4 (0.012)	7.1 (0.008)	10.8 (0.001)	2.9 (0.090)	18.8 (0.000)	15.5 (0.000)	11.1 (0.001)	0.8 (0.384)	5.2 (0.023)	7
Conditional t	2.4 (0.119)	2.6 (0.110)	2.8 (0.093)	0.4 (0.540)	0.3 (0.564)	3.1 (0.078)	2.9 (0.087)	2.6 (0.107)	0.2 (0.630)	0
Conditional EVT (t)	0.3 (0.578)	2.3 (0.128)	2.3 (0.128)	0.3 (0.575)	1.3 (0.247)	2.2 (0.140)	2.1 (0.144)	0.3 (0.558)	0.5 (0.470)	0
Conditional normal	0.8 (0.384)	0.3 (0.564)	0.0 (0.939)	2.0 (0.155)	1.9 (0.170)	1.1 (0.299)	0.2 (0.638)	4.4 (0.036)	1.1 0.293)	1
Conditional EVT (n)	0.3 (0.590)	1.8 (0.180)	2.1 (0.148)	0.3 (0.588)	1.2 (0.275)	2.1 (0.150)	0.3 (0.582)	0.4 (0.535)	0.4 (0.513)	0
Historical	5.9 (0.015)	3.7 (0.054)	20.1 (0.000)	3.8 (0.052)	23.7 (0.000)	11.5 (0.001)	13.4 (0.000)	0.5 (0.480)	11.8 (0.001)	6

Table 4: Backtesting results from the independence test. Tabel entries show LR_{ind} statistics (χ^2 distributed with 1 degree of freedom), and associated p-value in parentheses. Rightmost column shows number of times (out of 9) the model is rejected at significance level 0.05.

Dataset	N225	S&P500	DAX	ERICB	ELUXB	HOLMB	WTI	GOLD	WHT	Times
Length of test	7088	7086	7088	6564	6564	6564	6174	6776	6802	rejected
0.95 Quantile										
Expected	354	354	354	328	328	328	309	339	340	
Unconditional EVT	449 (0.000)	452 (0.000)	448 (0.000)	394 (0.000)	400 (0.000)	351 (0.202)	343 (0.008)	391 (0.004)	409 (0.000)	8
Conditional t	498 (0.000)	448 (0.000)	520 (0.000)	348 (0.257)	346 (0.308)	281 (0.007)	384 (0.000)	379 (0.028)	346 (0.738)	6
Conditional EVT (t)	391 (0.047)	375 (0.264)	392 (0.044)	353 (0.165)	369 (0.023)	334 (0.734)	319 (0.211)	344 (0.759)	378 (0.037)	3
Conditional normal	422 (0.000)	372 (0.340)	396 (0.025)	281 (0.007)	294 (0.054)	287 (0.019)	327 (0.085)	302 (0.042)	260 (0.000)	6
Conditional EVT (n)	393 (0.039)	366 (0.531)	376 (0.241)	350 (0.213)	373 (0.013)	330 (0.910)	326 (0.096)	331 (0.696)	363 (0.201)	2
Historical	428 (0.000)	419 (0.001)	439 (0.000)	380 (0.004)	384 (0.002)	348 (0.257)	321 (0.171)	342 (0.845)	393 (0.004)	6
0.99 Quantile										
Expected	71	71	71	66	66	66	62	68	68	
Unconditional EVT	137 (0.000)	169 (0.000)	131 (0.000)	93 (0.082)	105 (0.000)	85 (0.021)	83 (0.004)	142 (0.000)	109 (0.000)	8
Conditional t	81 (0.232)	83 (0.152)	73 (0.765)	66 (0.950)	57 (0.320)	52 (0.094)	42 (0.022)	61 (0.463)	67 (0.951)	1
Conditional EVT (t)	71 (0.952)	74 (0.676)	74 (0.676)	67 (0.852)	70 (0.576)	55 (0.214)	50 (0.241)	62 (0.541)	76 (0.329)	0
Conditional normal	135 (0.000)	142 (0.000)	124 (0.000)	115 (0.000)	92 (0.002)	106 (0.000)	94 (0.000)	99 (0.000)	95 (0.002)	9
Conditional EVT (n)	71 (0.952)	80 (0.282)	72 (0.858)	73 (0.352)	70 (0.576)	66 (0.950)	54 (0.515)	60 (0.392)	70 (0.807)	0
Historical	100 (0.001)	124 (0.000)	114 (0.000)	84 (0.025)	99 (0.000)	93 (0.001)	84 (0.003)	81 (0.112)	94 (0.003)	8
0.995 Quantile										
Expected	35	35	35	33	33	33	31	34	34	
Unconditional EVT	96 (0.000)	139 (0.000)	90 (0.000)	55 (0.079)	60 (0.000)	55 (0.000)	51 (0.000)	109 (0.000)	63 (0.000)	8
Conditional t	30 (0.400)	29 (0.312)	27 (0.177)	35 (0.662)	33 (0.930)	24 (0.136)	24 (0.357)	28 (0.344)	28 (0.344)	0
Conditional EVT (t)	33 (0.800)	31 (0.501)	31 (0.501)	32 (1.000)	40 (0.219)	31 (0.861)	30 (0.927)	34 (0.931)	42 (0.168)	0
Conditional normal	85 (0.000)	96 (0.000)	81 (0.000)	81 (0.000)	65 (0.000)	76 (0.000)	60 (0.000)	68 (0.000)	61 (0.000)	9
Conditional EVT (n)	32 (0.673)	36 (0.866)	33 (0.800)	31 (0.861)	42 (0.114)	32 (1.000)	30 (0.927)	36 (0.667)	38 (0.491)	0
Historical	60 (0.000)	75 (0.000)	59 (0.000)	48 (0.011)	49 (0.008)	52 (0.002)	44 (0.013)	41 (0.227)	52 (0.003)	8

Table 5: Backtesting results from the exact binomial test. Number of VaR breaks for each model, asset, and confidence level is shown. p-values from an exact two-sided binomial test are shown in parentheses. Rightmost column shows number of times (out of 9) the model is rejected at significance level 0.05.

5.1.1 Chrisoffersen's tests

We begin by summarizing the results from the Conditional Coverage test, shown in table 2:

- We first note that the two unconditional approaches, historical simulation and unconditional EVT, perform very poorly at all confidence levels. These models can be rejected at the 0.001 significance level in all cases except one.

- Second, we note that the conditional normal model performs poorly as well, particularly at the two higher confidence levels for the VaR.

- The conditionally t-distributed model performs very well at the 99.5% confidence level, fairly well at the 99% confidence level, but poorly at the 95% confidence level.

- The two conditional EVT models perform best overall, particularly at the higher confidence levels.

These comments also seem to hold for the unconditional coverage test; see table 3. The independence tests (results shown in table 4, show that all four conditional models perform well, particularly at the higher confidence levels. The unconditional EVT and the HS models perform badly, as expected.

5.1.2 Exact binomial test

The results of the exact binomial test should be close to those of the unconditional coverage test. They may be summarized as follows:

- The results indicate that the two unconditional approaches, historical simulation and unconditional EVT, perform very poorly at all confidence levels (for the VaR).

- The conditional normal model performs badly overall, but especially so at the 99% and 99.5% confidence levels.

- The conditional t-distributed model performs very well at the two higher confidence levels, but is just as bad as the conditional normal model at the 95% confidence level.

- The conditional EVT model with normally distributed residuals performs best overall, with no rejections (for a significance level of 0.05) at the two highest confidence levels, and two rejections at the 95% confidence levels.

- Finally, the conditional EVT model with t-distributed residuals performs very well at high confidence levels. It is almost as accurate as the conditional EVT model with normally distributed residuals at the 95% confidence level, with only one rejection more.

All in all, the results indicate that the conditional EVT models perform best. Conditional models in general perform better than unconditional models, but the conditional normal model fails often at high confidence levels for the VaR.

6 Conclusion

This thesis examined and compared the efficency of four models for calculating Value at Risk: an AR(1)-GARCH(1,1) model for which the innovations were assumed to be normally or t-distributed, and the same model, but for which the distribution of the tails of the innovations were modelled with the Peak-Over-Threshold method from Extreme Value Theory. For further comparison, an unconditional EVT model and a basic historical simulation model were included. The models were applied to nine data sets: three large stock indices, three blue chip stocks, and three commodities. We evaluated each model by employing two backtesting procedures: Christoffersen's backtest and an exact binomial test. The results indicate that the AR(1)-GARCH(1,1) model, for which the VaR-quantile is estimated by fitting a Peak-Over-Threshold model to the standardized residuals, is superior to the unaltered AR(1)-GARCH(1,1) model, especially at high confidence levels for the VaR-estimates. Furthermore, we find that the historical simulation and the unconditional extreme value theory models perform very poorly. These findings are similar to those of McNeil & Frey (2000), although we improved on their backtesting methodology by also testing for independence of the VaR-breaks.

We recognize that these conclusions are not to be interpreted as general facts, but may still count as evidence toward the validity of conditional EVT approaches to modelling Value at Risk. Further studies may wish to investigate closer the difference in accuracy of these types of models when applied to different asset classes, time periods, and market conditions. With this essay, we hope that the reader has gained some insight into the topics Value at Risk and Extreme Value Theory, and that he or she has been inspired to further study the fascinating field of risk management and the statistical theory of extremes, as we have.

References

Alexander, C. (2008), *Market risk analysis: Value-at-risk models*, John Wiley & Sons.

Artzner, P., Delbaen, F., Eber, J.-M. & Heath, D. (1999), 'Coherent measures of risk', *Journal of Economic Perspectives* Vol. 15 (no. 4), pages 157–168.

Balkema, A. A. & de Haan, L. (1974), 'Residual life time at great age', *The Annals of Probability* Vol. 2 (no. 5), pages 792–804.

Basel Committee on Banking Supervision (2010), Sound practices for backtesting counterparty credit risk models, Technical report, Bank for International Settlements.

Bekiros, S. D. & Georgoutsos, D. A. (2005), 'Estimation of value-at-risk by extreme value and conventional methods: a comparative evaluation of their predictive performance', *Journal of International Financial Markets, Institutions and Money* Vol. 15 (no. 3), pages 209–228.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroscedasticity', *Journal of Econometrics* Vol. 31 (no. 3), pages 307–327.

Byström, H. N. E. (2004), 'Managing extreme risks in tranquil and volatile markets using conditional extreme value theory', *International Review of Financial Analysis* Vol. 13, pages 133–152.

Christoffersen, P. F. (1998), 'Evaluating interval forecasts', *International Economic Review* Vol. 39 (no. 4), pages 841–862. Symposium on Forecasting and Empirical Methods in Macroeconomics and Finance.

Christoffersen, P. F. (2003), *Elements of financial risk management*, Academic Press.

Coles, S. (2001), *An introduction to statistical modeling of extreme values*, Springer-Verlag London Ltd.

Daníelsson, J. (2011), *Financial Risk Forecasting: The Theory and Practice of Forecasting Market Risk with Implementation in R and Matlab*, (The Wiley Finance Series), John Wiley & Sons.

Daníelsson, J. & de Vries, C. G. (2000), Value-at-risk and extreme returns, Technical report.

Daníelsson, J., Samorodnitsky, G., Sarma, M., Jorgensen, B. N. & de Vries, C. G. (2005), Subadditivity re-examined: the case for value-at-risk, FMG discussion papers, London School of Economics.

de Haan, L. & Ferreira, A. (2006), *Extreme value theory: an introduction*, Springer.

Dowd, K. (2005), *Measuring market risk*, second edn, John Wiley & Sons.

Einhorn, D. & Brown, A. (2008), 'Private profits and socialized risk', *GARP Risk Review* pages 10–26.

Embrechts, P., Klüppelberg, C. & Mikosch, T. (1997), *Modelling extremal events for insurance and finance*, Springer.

Engle, R. F. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation', *Econometrica* Vol. 50 (no. 4), pages 987–1007.

Engle, R. F. (2001), 'GARCH 101: The use of ARCH/GARCH models in applied econometrics', *Mathematical Finance* Vol. 9 (no. 3), pages 203–228.

Fama, E. F. (1965), 'The behavior of stock-market prices', *The Journal of Business* Vol. 38 (no. 1), pages 34–105.

Gilli, M. & Këllezi, E. (2006), 'An application of extreme value theory for measuring financial risk', *Computational Economics* Vol. 27 (no. 2-3), pages 207–228.

Jorion, P. (2001), *Value at risk: the new benchmark for managing financial risk*, second edn, McGraw-Hill.

Kotz, S. & Nadarajah, S. (2000), *Extreme value distributions: theory and applications*, Imperial College Press.

Kuester, K., Mittnik, S. & Paolella, M. S. (2006), 'Value-at-risk prediction: A comparison of alternative strategies', *Journal of Financial Econometrics* Vol. 4 (no. 1), pages 53–89.

Mandelbrot, B. (1963), 'The variation of certain speculative prices', *The Journal of Business* Vol. 36 (no. 4), pages 394–419.

McNeil, A. J. & Frey, R. (2000), 'Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach', *Journal of Empirical Finance* Vol. 7 (no. 3-4), pages 271–300.

McNeil, A. J. & Saladin, T. (1997), The peaks over thresholds method for estimating high quantiles of loss distributions, *in* 'Proceedings of the XXVIIth international Astin Colloquium', pages 23–43.

McNeil, A. J., Frey, R. & Embrechts, P. (2005), *Quantitative risk management: concepts, techniques, and tools*, Princeton University Press.

Ozun, A., Cifter, A. & Yilmazer, S. (2010), 'Filtered extreme-value theory for value-at-risk estimation: evidence from turkey', *The Journal of Risk Finance* Vol. 11 (no. 2), pages 164–179.

Pickands, III, J. (1975), 'Statistical inference using extreme order statistics', *The Annals of Statistics* Vol. 3 (no. 1), pages 119–131.

Ribatet, M. (2011), *POT: Generalized Pareto Distribution and Peaks Over Threshold*. R package version 1.1-1.

Tolikas, K., Koulakiotis, A. & Brown, R. A. (2007), 'Extreme risk and value-at-risk in the German stock market', *The European Journal of Finance* Vol. 13 (no. 4), pages 373–395.

Tsay, R. (2010), *Analysis of financial time series*, third edn, John Wiley & Sons.

U.S. Securities and Exchange Commission (n.d.), 'Market indices', http://www.sec.gov/answers/indices.htm.

Wickham, H. (2009), *ggplot2: elegant graphics for data analysis*, Springer New York.

Wuertz, D. & Chalabi, Y. (2010), *timeSeries: Rmetrics - Financial Time Series Objects*. R package version 2130.90.

Wuertz, D., with contribution from Michal Miklovic, Y. C., Boudt, C., Chausse, P. & others (2009), *fGarch: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling*. R package version 2110.80.

Yamai, Y. & Yoshiba, T. (2002), 'Comparative analyses of expected shortfall and value-at-risk: Their estimation error, decomposition, and optimization', *Monetary and Economic Studies* Vol. 20 (no. 1), pages 87–122.