Non-normality in Financial Markets and the Measurement of Risk

Dissertation

zur Erlangung des Grades

Doktor der Wirtschaftswissenschaft (Dr. rer. pol.)

der Juristischen und Wirtschaftswissenschaftlichen Fakultät
der Martin-Luther-Universität Halle-Wittenberg

vorgelegt von
Diplom-Ökonom
Christian Lau

Halle (Saale),

September 2014

1. Gutachter: Prof. Dr. Jörg Laitenberger

2. Gutachter: Prof. Dr. Claudia Becker

Tag der Verteidigung: 23.06.2015

Für Anneliese und Rupert Lau.

"Essentially,
all models are wrong,
but some are useful."
Box and Draper (1987, p. 424)

Contents

List of Figures

List of Tables

1 Introduction

One of the most important mandates for a financial company is computing and managing risk. However, as outlined, for example, by Christoffersen (2003) and McNeil et al. (2005), portfolio theory postulates that, based on the investor's appetite for risk, risk associated with a certain asset can be eliminated by investing in a diversified portfolio; therefore, this kind of risk taking will not yield excess-returns. Investors rather should buy a combination of risk-free asset and market portfolio. Consequently, companies need not be concerned about risk management, because the investor chooses the level of risk. This naturally raises the question whether companies should be interested in managing risk at all?

Indeed, in *perfect markets*[1], risk management becomes irrelevant (cf. Fite and Pfleiderer 1995). However, in reality, *perfect markets* do not exist and companies have to pay attention to risk management. In this context, the most important factors are *bankruptcy costs*, *taxes*, and *cost of capital*: A bankruptcy includes several costs such as those for lawyers or the closure of the company. Even a rumor of possible bankruptcy can diminish business prospects. Tax systems allow losses to be carried forward to offset future earnings. If we reduce the volatility of future cashflows, this will lower the future tax payment's net present value – and the present value of the company increases. Risk is also introduced when there is asymmetrical information between companies and investors, in which case raising capital may be more expensive. Typically, this is the case after a company incurs losses and urgently needs new money. In summary, risk management is necessary because it can lower the probability of bankruptcy, reduce tax payments, or prevent liquidity gaps. If we consequently agree in managing risk, we need to assess risk first. Accordingly, this thesis is focused on the measurement of risk.

The remainder of the introduction is structured as follows. The next section provides a brief introduction to risk and an approach to measure risk: *value at risk* (VaR). Different ways of computing VaR are examined with respect to the stylized facts of financial time series. This section then discusses multivariate extensions as well as drawbacks of the VaR. The second section outlines the objectives of this thesis, which includes

[1] A perfect market implies equal trading conditions for every market participant, no transaction costs, and no information asymmetry.

the identification of academic voids. The final section points out the links between the three principal self-contained chapters, summarizes each chapter, and concludes with key findings of this thesis and suggestions for further research.

1.1 Measuring Risk with Value at Risk

Subsection 1.1.1 outlines the different kinds of risk discussed in the literature. Subsection 1.1.2 introduces the risk measure VaR. In Subsection 1.1.3, stylized facts of financial time series are explained. Based on these, three general approaches to compute VaR are presented in Subsection 1.1.4. Subsection 1.1.5 is focused on a portfolio point-of-view with respect to VaR. Drawbacks and alternatives to the VaR concept are discussed in Subsection 1.1.6.

1.1.1 Types of Risk

Economic literature discriminates between four main kinds of risk (cf. Jorion 2007). *Market risk* is concerned with possible price movements in financial markets that cause losses in the participant's portfolio. Depending on the underlying risk factor, market risk can be subdivided further into several categories including equity, interest, and commodity risk (cf. Dowd 2005). This thesis is largely concerned with market risk because its management can be regarded as a decisive factor for a financial institution's success (cf. Cremers et al. 2012). *Liquidity risk* occurs when financial items cannot be sold without allowing a discount for the counterpart. *Credit risk* includes a too-late redemption of debt, a partial or full default or a deterioration of the rating. *Operational risk* involves "the risk of direct or indirect loss resulting from inadequate or failed internal processes, people and systems or from external events." (Basel Committee 2001, p. 2).[2] Financial companies have to provide equity in order to shoulder unexpected losses resulting from those risks. These regulatory requirements are fixed in Basel II/III (banks) and Solvency II (insurances), respectively. For this purpose, obviously, the amount of risk has to be measured.

1.1.2 Value at Risk

Because of its easy computation, a self-evident measure of risk is the variance. We find its application in seminal work, for example, portfolio theory (Markowitz 1952) or option pricing (Black and Scholes 1973). The use of variance is feasible as long as we can apply Gaussian distribution because the latter is completely described by the first and second moment, that is, mean and variance. As we will see later, this point of view becomes very limited because, in the case of financial returns, models incorporating

[2] In addition, *legal risks* include, for example, unexpected changes in law.

higher moments (skewness and kurtosis) are more appropriate. Furthermore, whereas the variance considers both positive and negative deviations, we are mainly interested in the latter in terms of risk management (cf. Wagner 2000).

The main key figure used to measure (market) risk is the VaR. Assuming a certain level of confidence $1-\theta$, according to Christoffersen (2003), the $VaR(\theta)_{t+1}$ of an asset is defined as the number such that the probability that the asset's next day return r_{t+1} will be below $-VaR(\theta)_{t+1}$, is θ: $Pr(r_{t+1} < -VaR(\theta)_{t+1}|\Omega_t) = \theta$, where Ω_t is the information set available in t. From a statistical point of view, the VaR simply corresponds to the quantile of a distribution of returns (cf. McNeil et al. 2005). The VaR, in the broad sense, gives the worst loss that will not be exceeded with a certain probability over a fixed target horizon (cf. Jorion 2007). Alexander (2010), Dowd (2005), Holton (2003), or especially, Jorion (2007), give a thorough treatment of the computational methods, applications and drawbacks of VaR. Holton (2002) outlines that the use of VaR can be traced back to 1922 when the NYSE asked their members for capital requirements. In the 1980's, financial institutions began to apply proprietary VaR measures. J.P. Morgan and Reuters (1996) used VaR in a more professional way (RiskMetrics) and offered it as a service to other companies.

Today, VaR is a widely accepted standard for calculating financial risk. The Basel Committee on Banking Supervision adopted the VaR in their Basel regulations for banks (cf. Basel Committee 2006, 2011). Furthermore, VaR can be used as a risk limit for traders or managers because, in comparison to a stop-loss limit[3], the advantage of VaR limits is given by its prospective nature (cf. Holton 2003). In addition, VaR is a welcome reporting tool for senior management and shareholders. This thesis discusses applications of VaR in different financial markets.

1.1.3 Stylized Facts of Financial Time Series

The simplest way to calculate VaR is by using a Gaussian assumption for the return distribution. Although this approach is very convenient – not least because of an easily available extension to the multivariate case – it has serious drawbacks. Academic literature going back to Mandelbrot (1963) provides strong empirical evidence that Gaussian distributions cannot capture appropriately the behavior of financial returns. This problem is magnified particularly when we are dealing with the tails of distributions such as in a VaR computation.

The contradictoriness between Gaussian assumption and realized return distribution is the consequence of the (empirically) observed *stylized facts* of financial time series (see Cont (2001) for a complete list). Stylized facts are (statistical) properties of (financial) time series, evidenced across a wide range of markets, time periods and instruments. In

[3] A stop-loss limit represents that amount of money that should not be exceeded by an asset's or portfolio's loss.

accordance with Christoffersen (2003), we only discuss the most relevant stylized facts of daily financial returns for which we have to account when computing VaR.

i. The unconditional distribution of financial returns is not Gaussian, because financial returns display *heavy tails* (sometimes referred to as *fat tails* or *excess kurtosis*). This implies that extreme returns will occur more frequently than predicted by Gaussian distribution. Furthermore, the empirical return distribution is much more peaked around the mean and this shape is referred to as *leptokurtic*.

ii. Financial returns tend to be asymmetric or rather negatively skewed because of extreme drops and a lack of equally large positive movements.

iii. There is almost no autocorrelation in financial returns, thus we cannot predict future returns from their own past.

iv. The variance of financial returns shows positive correlation with its own lagged values. This appears as *volatility clusters* in figures showing the returns. Even after the standardization of returns by a time-varying measure of volatility, returns still display heavy tails.

v. Correlation is also time-varying. In periods of crisis, correlation between assets tends to be greater compared to regular periods.

1.1.4 Approaches to Calculate VaR

In terms of risk management, we should apply a VaR calculation method that accounts for the stylized facts if we want to receive reliable values. There exists a great body of literature dealing with comparisons of different VaR calculation techniques (cf., e.g., Abad and Benito 2013; Rachev et al. 2010; Marinelli et al. 2007; Kuester et al. 2006; Pérignon and Smith 2008). We can break down the approaches to calculate VaR into three categories.

The first option is to use an *analytical* approach. Here, a model is required that reproduces the autoregressive properties of the time series. The most simple and viable technique is an exponentially weighted moving average (EWMA). An EWMA applies weighting factors to each element of a time series while the weighting factors decrease exponentially as we move back in time. This idea is adopted as a model for the variance process by J.P. Morgan and Reuters (1996) and included in a set of risk measurement methods (RiskMetrics). In this case, tomorrow's variance is the sum of today's variance multiplied by λ and today's squared return multiplied by $1 - \lambda$. J.P. Morgan and Reuters (1996) conduct a large study to find the best choice of the decay factor λ. They compare the performance of different decay factors on the basis of the root mean square error and find that choosing $\lambda = 0.94$ leads to the most accurate volatility forecast. However, because volatility shows to have a long-run average, which is ignored by the EWMA, a broadly accepted and applied way is to model the variance with a GARCH model (Bollerslev 1986). In contrast to the EWMA, GARCH models include an innovation process that can be modeled with a distribution exhibiting skewness and kurtosis.[4] Referring to this, Bollerslev (1987) is the first to relax the assumption of conditional normality and instead assumes that the standardized innovation process follows a standardized Student's t-distribution. Further applications can be found, for example, in Forsberg and Bollerslev (2002), who introduce a GARCH process with normal inverse Gaussian innovations, or in Kim et al. (2010), who apply the tempered stable distribution. In general, parameter estimation is done by applying maximum likelihood estimation (MLE). Alternatively, estimation can be achieved in two steps by conducting a quasi MLE (QMLE). In this case, autocorrelation is removed first from the time series and a distribution under consideration is used to fit the standardized residuals. The most prominent contribution to this method is given by McNeil and Frey (2000), who focus only on the tails (rather than the whole distribution) because this is the important region in a VaR computation. Therefore, the authors propose to fix a threshold u and describe all standardized residual losses exceeding u (they belong to the distribution's tail) with a generalized Pareto distribution. The best choice for u is still in question (see Scarrott and MacDonald (2012) for a survey), therefore, we follow the rule of thumb of DuMouchel (1983) and choose u such that 10% of the whole sample belong to the tail. However, scientists face the problem of consistency when using QMLE (cf. Ling and McAleer 2003; McAleer et al. 2007; Shephard 1996).

The second approach to calculating VaR merges *simulation* techniques. A Monte Carlo simulation (MCS) is based on random numbers and uses, for example, GARCH-type models to produce a large number N of hypothetical returns. The VaR is simply the $\theta \cdot N$ smallest value of the ordered returns. A thorough introduction to MCS with financial applications can be found in Glasserman (2004). The historical simulation (HS) is based solely on the empirical distribution of the real returns generated up to the actual date. Comparable with the MCS, given a number of observations N, the $\theta \cdot N$ smallest value of the ordered returns represents the VaR. In contrast to MCS and the analytical approach, no assumption about the underlying distribution has to be stated. However, as argued, for example, in Pritsker (2006), a standard HS assigns an equal weight of probability to each return. This implicitly assumes that the historical returns are iid, which is unfeasible, because the volatility level of assets tends to change over time (see stylized fact (iv)). As a hybrid approach, the powerful filtered historical simulation (FHS) (cf. Hull and White 1998; Barone-Adesi et al. 1999) overcomes this drawback and combines both GARCH framework and HS to account for changes in

[4] In the last several decades, a large number of GARCH based models have been put forward. Bollerslev (2009) provides a glossary for different GARCH models. A very general overview of various classes of volatility models can be found, for example, in Andersen et al. (2006). A detailed description of the GARCH model and two popular extensions is given in Section 3.2.

the volatility level. Kuester et al. (2006), for example, show that FHSs are good models for VaR calculation. In contrast to the MCS, the number of returns N is limited in case of the (F)HS. When conducting VaR applications subsampling procedures are required. No ultimate solution concerning the optimal length of the subsample (in some applications also referred to as the window length) exists. As outlined by Christoffersen (2003), recommendations vary between 250 days (that corresponds to one year) and 1000 days (4 years). Kuester et al. (2006) also note that for smaller sample sizes, calculating VaR should become more challenging. However, even shorter and longer time periods are of interest. Hendricks (1996), for example, computes VaR using the HS as well as a parametric approach for window lengths between 50 and 1250 days. He finds that for greater subsamples the VaR becomes more stable.

The third approach is to ignore the original return series and instead model the quantile (the VaR) directly in an autoregressive specification. Engle and Manganelli (2004) introduce this approach as conditional autoregressive VaR (CaViaR), and although an elegant approach, results are mixed. Kuester et al. (2006) also find that CaViaR models are substandard. In contrast, Lima and de Néri (2007) apply a quantile regression with ARCH effect and find a superior performance compared to techniques requiring distributional assumptions.

1.1.5 Portfolio Point-of-View

So far, we have ignored the fifth stylized fact – the time-varying correlation. When we examine a portfolio of assets, they will show interdependencies between each other. These interdependencies might be rather high (in times of crisis) or low (in normal times). Thus, considering portfolio risk, practitioners might ask for a multivariate framework to assess downside risk instead of a univariate one. We briefly look at three alternatives to evaluate portfolio risk.

The first way is to assume a multivariate normal distribution. Although this method is often used in practice because of its simplicity, the other stylized facts of financial returns are ignored. A second, broadly accepted way to evaluate risk is through the implementation of copula methods (a copula function builds a bridge between the marginal probability distributions of random variables and their joint distribution (see Cherubini et al. (2004) for applications of copula models in finance)). Applying Sklar (1959) (see McNeil et al. (2005) for an interpretation), it is possible to split risk modeling into two consecutive steps when using a copula: the modeling of the single risk itself and the choice of a suitable dependency structure of the single risk (the copula). Keeping this in mind, univariate modelling can be seen as a first step of modelling portfolio risk. The third way to evaluate risk is by using a multivariate GARCH-type model. This provides a time-varying structure for the variance, as in the univariate case, and allows assets to have dynamic correlations. Bauwens et al. (2006), for example, give a survey on multivariate GARCH models. The dynamic conditional correlation (DCC) model of Engle (2002a) is probably the most popular multivariate model in academia because it offers a reasonable trade-off between adequacy and obstacles in the estimation process (cf. McNeil et al. 2005).

Although we find elaborate approaches to model dependency in returns, there is, as yet, no unanimity whether this portfolio point of view is desirable, or whether a univariate analysis of a simple aggregation of the associated assets returns is sufficient (cf. Berkowitz and O'Brien 2002). The type of dependency structure that should be applied remains questionable. Furthermore, in the case of a large portfolio, the performance of the model may be substandard. However, building a multivariate model is generally a second step in the process of modelling financial times series. Consider, for example, the GARCH model – more than a decade passed between the publication of the GARCH model and its multivariate expansion. That is, when we consider new areas of research it is reasonable to first investigate single time series. This thesis attends to the univariate case and leaves expansions for multivariate cases for future work.

1.1.6 Drawbacks and Alternatives

Indeed, the VaR concept is very easy to understand. However, Artzner (1999) points out that VaR is not always a coherent risk measure, because it could lack in sub additivity. Sub additivity demands the risk measure of a portfolio of assets to be smaller than the sum over the risk measures of every single asset. This implies that diversification should reduce risk. Second, VaR gives no idea *how bad it really could be*. Correspondingly, Taleb (2007, p. 161) advises "Don't cross a river if it is four feet deep on average". Because VaR just gives a lower bound of a possible loss, no expectation formation in terms of a concrete value of an expected loss is involved in this risk measure.

In a consultative document, the Basel Committee (2012) suggests that we consider *expected shortfall* (ES) as an alternative to VaR to measure downside risk. ES, sometimes termed *expected tail loss*, *conditional VaR*, or *average VaR*, is defined as the expected value of the loss given that VaR is exceeded: $ES_{t+1} = -E_t[r_{t+1}|r_{t+1} < -VaR]$, where $E_t[.]$ means taking the expectation with all information available up to and including day t (cf. Christoffersen 2003). In contrast to VaR, ES returns a more concrete assessment of the real risk of an asset. Furthermore, it can be shown that ES is a coherent risk measure in terms of sub additivity (cf. Artzner 1999). However, because of its conditionality, ES is not as comprehensive as VaR, and far less common. Hence, this thesis focusses on VaR.

1.2 Objective of the Thesis

The objective of this thesis is to provide new insights on risk associated with financial markets. Companies provide equity to shoulder unexpected losses arising from risks, therefore, an accurate assessment of this risk is maybe the most important challenge in risk management.

Because the sovereign debt crisis led to totally new challenges, a contemporary investigation of the European government bond market is essential. Although we know many of the characteristics of risk stemming from, for example, stock markets, less is known about risk arising from bond markets. The investigation of the latter could lead to fundamental implications for interest rate modeling, fixed income portfolio management, and monetary policy making. So far, risk management concerning bonds focuses on key aspects such as duration and convexity[5]. The lack of knowledge is striking considering the fact that bond markets exceed stock markets in terms of capitalization (cf. Laopodis 2008). This present thesis attempts to close this knowledge gap and, by elaborating a new approach, helps convey a better understanding of risk associated with the European government bond market. As outlined in stylized fact (i) financial returns show a non-Gaussian behavior. Therefore, we test the Gaussian assumption first and determine if the distribution of European government bond returns exhibits higher moments. Based on those findings, we assess distributions matching the characteristics of European government bond returns more appropriately to compute the risk in a VaR study. Then, we are able to give an answer to the question "What is the risk of an investment in European government bonds?".

The second market under consideration is the electricity forward market. Although electricity itself belongs to the group of energy commodities, electricity forwards are settled in cash and, therefore, can be treated as financial risk. This market has some very special and interesting characteristics, such as the non-storable nature of the underlying product or the lack of an analytical connection between spot and forward price. Electricity forward markets are a relatively new field of research and the scarce existing research is mainly based on synthetic forwards. In contrast, we want to examine the risk of an investment in real electricity forwards, that is, we wish to calculate the risk an investor actually faces when trading an electricity forward. The visual inspection of electricity forward return time series suggests the use of an autoregressive framework for both mean and variance. Again, we compute the risk associated with an investment in electricity forwards with a VaR calculation.

Third, this thesis provides a new way of calculating VaR in a simple but comprehensive way. Although we are mainly interested in a VaR that is high enough to shoulder unexpected losses, the VaR should still be as low as possible, thus minimizing the

[5] Duration is the sensitivity of the bond's price with regard to the interest; convexity is the second relative derivation of the bond's price with regard to the interest.

amount of capital that has to be allocated. The solution to this catch-22 has a tremendous impact on the calculation of risk and, therefore, risk management: if a company is able to achieve both goals simultaneously, it gains a competitive advantage, since it is able to reallocate equity to underwrite further risks. There are many sophisticated ways to assess financial risk, but none that is concerned with this particular issue. Moreover, it is still challenging to calculate VaR in volatile periods as during the financial crises. Therefore, the third objective of this thesis is to define a new model for computing risk that considers both issues. This is achieved by combining three fields of research. The new model (a) accounts for the stylized facts of financial time series, (b) includes the evaluation of intraday information, and (c) uses method of moments estimation for the parametrization of a suitable distribution. Data from the stock market is used for the empirical study because this market provides dependable and easily available data.

1.3 Outline of the Thesis

The three principal chapters in this thesis show a high degree of connection. With regard to contents, all three chapters assess risk in different financial markets. Furthermore, the chapters are similar in their way of proceeding. First, as outlined in Subsection 1.1.5, this thesis is concerned with the univariate case. Initially, each chapter is based on a thorough analysis of the relevant time series. Beside the computation of higher moments, this includes, for example, an examination of the Lilliefors test (Lilliefors 1967) and the Jarque-Bera test (Bera and Jarque 1987): these are two of the prominent tests for assessing normality (cf. Cottin and Döhler 2009). In each case, the analysis shows that normality has to be rejected and higher moments causing heavy tails are an important feature of the respective markets. Referring to this stylized fact, we, secondly, try to find distributions matching the characteristics of the unique markets. Third, the chapters close with a VaR computation to calculate the respective risk. In this context, a backtesting methodology as outlined by Christoffersen (2003) is conducted to assess VaR violations concerning the correct frequency and independence. According to the stylized facts of the financial time series and depending on the research available from the literature, the chapters naturally differ in the modelling method.

Research on the bond market, to date, is relatively restricted. Therefore, the first chapter of this thesis starts with an unconditional modelling of European government bond returns. Subsequently, we concentrate on the first and second stylized facts (heavy tails and skewness), while ignoring the fourth one, in particular, and test the appropriateness of three distributions fitted with the maximum likelihood method.[6] We use the most convincing distribution to calculate VaR with an (unconditional) analytical

[6] In Chapter 4, the normal inverse Gaussian is used to model a financial time series. We abstain from applying this distribution here, because there exists no reliable MLE (cf. Karlis 2002).

approach. The third chapter conducts analytical as well as simulation approaches to calculate VaR for electricity forward returns. Concerning the modelling of electricity forwards, the literature agrees in accounting for heavy tails and conditional volatility modelling. However, contrary to the third stylized fact (no autocorrelation in the returns), we find evidence for autocorrelation in the returns. In contrast to bond and electricity markets, the amount of literature about risk in stock markets is overwhelming. We extend the fourth stylized fact by allowing higher order moments to have a time varying property, too. Since these conditional higher moments are not a new issue to finance literature, we use intraday data for computation to deliver erudite forecasts for the recent moments and consequently present a new approach to calculate VaR. The following subsections provide summaries of the three chapters; the final subsection summarizes the key findings and presents directions for future research.

1.3.1 On the Distribution of Government Bond Returns: Evidence from the EMU

In contrast to stock markets, academic research neglects a comprehensive analysis of bond markets risk, which seems remarkable because bond markets exceed stock markets in terms of capitalization. In Chapter 2 "On the Distribution of Government Bond Returns: Evidence from the EMU", we attempt to remedy this situation by studying the returns of eleven European government bonds with maturities of one, five and ten years. Initially, we examine whether a Gaussian hypothesis is acceptable for European government bond returns. We present evidence for an overwhelming rejection of the Gaussian assumption because we find great excess kurtosis and considerable skewness in the data. Therefore, we apply three alternative distributions, all accounting for excess kurtosis and partially for skewness, to the data and find that the stable distribution offers the best overall fit. During the financial crisis, prices of government bonds began to behave differently, raising the question whether structural breaks occur in the time series. By applying a Quandt likelihood ratio test, we identify structural breaks in most of the time series and re-estimate the parameters of the distributions under consideration for the time after the structural break. This second analysis confirms our previous findings and shows that the stable distribution still offers the best fit. Next, we use our research results to point out implications for risk management. We compare VaR based on the alternative distributions and find that the stable distribution returns the best results in terms of backtesting.

1.3.2 Measuring Risk in Electricity Forward Returns

A broad area of research is focused on describing spot prices of electricity. In contrast, research about the derivative electricity market is rather scarce. In Chapter 3 "Measuring Risk in Electricity Forward Returns", we calculate and backtest different VaR methodologies for electricity forward contracts. In theory, there exists a strong relationship between a forward and its underlying (asset). Interestingly, this is not the case for electricity forwards, this means it is not possible to establish an analytical relationship as, for example, in the case of stocks. In contrast to electricity forward prices we find a high degree of seasonality and massive, sudden jumps in the corresponding underlying, that is, the electricity spot price. This study includes data of yearly and quarterly forward returns. Our first analysis reveals that the Gaussian hypothesis has to be rejected. Whereas there is only negligible skewness in the data, we determine a great excess kurtosis. Moreover, we find evidence for autocorrelation in the returns and squared returns and, therefore, propose a conditional modelling of both corresponding moments. A simple momentum strategy is demonstrated as a consequence of autocorrelation in the returns. Subsequently, we use an ARMA process and describe the variance with three different GARCH-type models[7]. In terms of the Bayesian information criterion, a simple GARCH(1,1)-t specification gives the best overall ex-post fit. This raises the question whether an autoregressive modelling of the mean is necessary. Next, we check the forecasting abilities of the best models in a VaR study. Our initial results are not confirmed and autocorrelation in returns indeed becomes an issue. On average, an ARMA(1,0)-GARCH(1,1)-t model included in an FHS provides the most reliable framework in terms of backtesting.

1.3.3 A simple NIG-type Approach to calculate Value at Risk based on Realized Moments

Surprisingly, although we find a great number of viable ways to compute VaR, practitioners prefer rather simple methods such as the Gaussian assumption or HS. Possible reasons for this include lack of comprehension or the occurrence of model risk. Chapter 3 "A simple NIG-type approach to calculate Value at Risk based on Realized Moments", therefore, presents a simple but comprehensive approach to calculate VaR even in volatile periods. First, the *realized variance* concept based on intraday data is expanded by realized skewness and kurtosis. We need accurate forecasts to calculate VaR for the next day, therefore, we derive forecasts of variance, skewness, and kurtosis with a simple EWMA. Next, we use these forecasts in a method of moments to parametrize a distribution that exhibits skewness and excess kurtosis. For this purpose, the normal inverse Gaussian (NIG) distribution is an obvious choice because it accounts for higher order moments and offers an explicit way to compute parameters with the method of moments. Once parameters are calculated, the quantile (the VaR) can be computed simply by applying the inverse function. Using intraday data from the DAX, we com-

[7]To avoid redundancy, please note that a detailed description of those models is given in Chapter 3.2 instead of here.

pare the results of this VaR methodology with other models based on realized variance and daily data on different confidence levels in terms of backtesting. Keeping in mind that VaR should be as high as necessary, but as low as possible, we identify a superior performance of the presented method. Although this method provides good results, some possible modifications as well as an expansion to a multivariate application are discussed and left for future research.

1.3.4 Key Findings and Future Research

Chapter 2 shows that using a distribution accounting for heavy tails in government bond returns is crucial when modelling those returns and calculating VaR. However, our analysis is unconditional because we do not consider a time-varying variance. With regard to stylized fact (iv), a further question is whether the application of models incorporating a time-varying variance helps to explain more accurately the risk of government bond returns. In Chapter 3, we find that an autoregressive framework for the mean as well as for the variance provides a feasible setting for calculating the risk of electricity forward returns in terms of VaR. Because electricity markets tend to show correlation across forwards (cf. Frestad et al. 2010; Solibakke 2010), the next step is the application of a multivariate framework. Based on intraday data and taking the example of a stock market, Chapter 4 shows that the NIG distribution parametrized by using conditional forecasts for variance, skewness, and kurtosis improves VaR calculation in such a way that the VaR is lower but still feasible in terms of backtesting. This finding may have a striking impact on the allocation of equity in financial companies. A very interesting feature of this proceeding is the expansion to a portfolio point of view – that is left for future research.

Please note that, in some instances, an approach or argument may be repeated. This redundancy is necessary to ensure that single chapters can be treated as self-contained.

2 On the Distribution of Government Bond Returns: Evidence from the EMU

This chapter is based on the corresponding article by Gabriel and Lau (2014).

2.1 Introduction

International bond markets clearly exceed equity markets in terms of capitalization (cf. Laopodis 2008). Thus, investigating these markets could have important implications for interest rate modeling, fixed income portfolio management, and monetary policy making. However, equity markets attract considerably more attention in the finance literature than do bond markets. The European Monetary Union (EMU) bond market is particularly unique in that it accommodates economies with different levels of credibility and fiscal discipline in one currency (cf. Beber et al. 2009).

The objective of this chapter is to investigate the statistical distribution of price changes in European government bonds. For the period 1999 to 2012, we investigate all countries that joined the EMU before 2001. We exclude Luxembourg from our analysis since its public debt market is negligible. The data frequency is daily bond returns with one-, three-, five- and ten-year maturity. Descriptive statistics and tests of normality lead to a clear rejection of the Gaussian assumption. We therefore propose alternative distributions and fit the Student's t, skewed Student's t, and stable distribution to the data. Since the Euro crisis leads to a shift in the mean and an increase in the standard deviation, we test each time series for a structural break and separately study the crisis period.[1] Finally, a value at risk (VaR) application contributes to better understanding the implications that can be derived from the distributional assumption.

The type of distribution of financial returns is an essential assumption for mean-variance portfolio theory, pricing of financial derivatives, and many other applications. Mandelbrot (1963) and Fama (1965) reject normality because heavy tails are a key feature of financial returns. They and other authors propose various distributions that account for excess kurtosis (cf. Press 1967; Praetz 1972; Blattberg and Gonedes 1974;

[1] It is important to note that our analysis is based on unconditional distributions. GARCH-type models (cf. Bollerslev 1986) are beyond the scope of the present chapter.

Peiró 1994). However, investors not only have an aversion to the second and fourth moment, but also a preference for positive first and third moments. Hence, skewness is important for modeling financial returns (cf. Kon 1984; Hansen 1994; Young and Graff 1995; Peiró 1999; Rachev and Mittnik 2000; Aparicio and Estrada 2001). This branch of the literature is mainly concerned with equity returns, whereas the EMU, which is the market of interest in this chapter, is more often discussed in debt capital market research.

Prior to the EMU, we observe converging yields and harmonizing prices of Euro-denominated government bonds (cf. Baele et al. 2004; Codogno et al. 2003; Hartmann et al. 2003). Decreasing government financing costs are one reason for the significant growth of the European bond market (cf. Pagano and von Thadden 2004). There is a large body of literature concerned with the European bond market and its interactions with other major bond markets (cf. Cappiello et al. 2003; Christiansen 2007; Abad et al. 2010). Laopodis (2008) conducts an extensive empirical study of the link between Euro and non-Euro government bonds for the period 1995 to 2006. However, he draws no conclusions as to which distribution fits the bonds' variation best. Rachev et al. (2003) are the only authors who study the distribution of U.S. corporate bond returns.[2]

The present chapter contributes to the literature by providing a comprehensive study of EMU bond return distributions. To the best of our knowledge, we are the first authors to analyze the daily bond returns of all EMU countries with one-, three-, five-, and ten-year maturity. We test alternative distributions, account for structural breaks in the time series, and offer an application for risk management.

The remainder of this chapter is organized as follows. Section 2.2 reports some descriptive statistics and tests the normality assumption. Section 2.3 presents the theory of the proposed distributions, and Section 2.4 shows the parameter estimation results. Section 2.5 reports the results of a Quandt likelihood ratio test to ascertain if there is a structural break and then takes another look at the Euro crisis period. Section 2.6 presents a VaR application for government bond returns. Section 2.7 concludes.

2.2 Data and Test of Normality

In terms of capitalization, debt markets clearly exceed equity markets (cf. Laopodis 2008). Additionally, the EMU bond market is unique in providing debt for countries with different levels of credibility and fiscal discipline in one currency (cf. Beber et al. 2009). Therefore, we study government bonds issued by EMU members.

[2] Rachev et al. (2003) fit the stable distribution to U.S. corporate bond indices. Interest rate risk, measured with duration, and credit default risk are the main risk-driving factors of bonds. Using indices leads to a clustering of duration and rating, resulting in a less than clear view of the bonds' risk.

The dataset consists of all countries that joined the EMU before 2001 with the exception of Luxembourg. Countries that joined the Eurozone much later are excluded to avoid studying time series of considerable different length. The daily zero bond returns are provided by Datastream. The empirical study starts in 1999, when exchange rates for prospective Euro members were fixed. The sample period is January 1, 1999 to November 30, 2012, resulting in 3,627 data points for each time series.[3] The cross-section of bond returns are fixed maturities of one, three, five, and ten years.

Next, we calculate the bonds' daily return. Let $y_t(\tau)$ be the yield of a bond at time t and τ its time to maturity. At each point in time, we take the yield of the previous day $y_{t-1}(\tau_1)$ multiplied by the initial time to maturity τ_1 (= one, three, five, and ten years) and subtract it from today's yield $y_t(\tau_2)$ multiplied by the remaining time to maturity τ_2 ($= \tau_1 - 1$ day).[4] The log return $r_t(\tau)$ at time t with maturity τ reads:

$$r_t(\tau) = exp(-\tau_2 y_t(\tau_2) + \tau_1 y_{t-1}(\tau_1)) - 1. \tag{2.1}$$

In this manner we calculate the log return for every bond at each point in time.

Table 2.1 summarizes some descriptive statistics of EMU bonds. The left panel of the table shows the mean, standard deviation, skewness, and kurtosis of government bond returns. We apply the Lilliefors and Jarque-Bera goodness-of-fit tests of normality and report the results in the right panel (cf. Peiró 1999; Aparicio and Estrada 2001). In Table 2.1 and henceforth, the order of the countries follows their exposure to sovereign risk.[5] The table shows that the mean is positive and close to zero for the daily bond returns. The only exception is Greece with slightly higher and negative returns, which we treat as a special case throughout the chapter.[6] It is evident that returns increase with time to maturity. This indicates a normal term structure for most of the time series. Analogously, bond risk is an increasing function of time to maturity. Although yields of the short end are more volatile, the exposure to interest rate risk is much higher for bonds with a longer time to maturity. Values range from a low of $0.405 \cdot 10^{-3}$ (Germany one year) to $17.416 \cdot 10^{-3}$ (Portugal ten years). The order of countries suggests that standard deviation increases with exposure to interest rate and sovereign risk.

The assumption of normality indicates that returns are symmetrically distributed, i.e., exhibit a skewness of zero. The table shows that countries with low sovereign risk tend to have very low skewness (Germany to Belgium), whereas countries with high

[3]The time series for Belgium and Greece start in 2001, resulting in 3,150 data points.

[4]We subtract two days for a public holiday and three days for a weekend.

[5]The new phenomenon of sovereign risk in European government bonds is important for interpreting the variation of returns (cf. Gomez-Puig 2009; Bernoth et al. 2004; Sgherri and Zoli 2009). We calculate the average spread of ten-year bonds for each country over ten-year German bonds, which we assume to be the reference.

[6]Due to the imminent default of Greece, results for Greek bonds are more extreme throughout the study. For the sake of readability, we document the results for Greece only if they provide new insight.

Table 2.1: Descriptive statistics of European government bond returns

	m	Mean $[10^3]$	Std $[10^3]$	Skew	Kurt	goodness of fit LF	JB
GER	1	0.111	0.405	0.699	18.461	0.080	36,410
	3	0.144	1.364	-0.109	5.408	0.044	883
	5	0.174	2.370	-0.199	4.708	0.043	464
	10	0.230	4.170	-0.021	5.191	0.043	725
NET	1	0.115	0.643	-0.011	22.512	0.120	57,519
	3	0.147	1.557	0.194	12.273	0.068	13,015
	5	0.175	2.403	-0.150	6.202	0.049	1,562
	10	0.224	6.017	-0.250	11.473	0.068	10,884
FIN	1	0.118	0.586	-0.067	17.188	0.110	30,414
	3	0.148	1.486	-0.297	5.969	0.047	1,384
	5	0.178	2.726	-0.509	19.933	0.064	43,475
	10	0.229	5.807	0.637	17.037	0.070	30,016
FRA	1	0.112	0.534	-0.206	14.684	0.105	20,651
	3	0.147	1.385	-0.054	7.315	0.053	2,814
	5	0.173	2.401	0.087	15.839	0.053	24,910
	10	0.210	3.981	-0.396	11.879	0.044	12,005
AUS	1	0.116	0.656	0.208	18.229	0.161	35,064
	3	0.152	1.336	0.041	12.719	0.148	14,271
	5	0.182	2.345	-0.209	11.000	0.137	9,696
	10	0.229	5.908	-0.165	9.552	0.133	6,503
BEL	1	0.119	0.677	-0.801	30.893	0.141	102,383
	3	0.169	1.667	0.114	15.826	0.069	21,584
	5	0.208	2.640	-0.093	10.664	0.059	7,708
	10	0.268	4.676	-0.151	7.960	0.056	3,238
SPA	1	0.123	1.040	-0.750	29.064	0.170	102,979
	3	0.139	2.105	1.843	45.238	0.171	271,597
	5	0.144	3.122	2.041	38.123	0.145	188,900
	10	0.122	5.599	0.291	12.896	0.106	14,846
ITA	1	0.127	0.744	0.547	34.413	0.166	149,266
	3	0.150	2.151	1.490	38.693	0.125	193,822
	5	0.162	3.410	1.894	41.167	0.107	222,254
	10	0.168	5.614	0.995	28.085	0.097	95,671
IRE	1	0.153	1.647	2.886	69.022	0.209	663,589
	3	0.177	4.180	3.991	93.992	0.200	1,260,538
	5	0.182	4.923	1.060	43.201	0.161	244,843
	10	0.110	11.848	-1.837	53.103	0.186	381,299
POR	1	0.162	2.936	-0.912	46.671	0.226	288,646
	3	0.170	6.339	-2.620	101.927	0.230	1,482,741
	5	0.154	8.073	-1.155	110.194	0.211	1,736,848
	10	0.101	17.416	-0.523	32.179	0.175	128,796
GRE	1	-0.631	29.724	-6.020	268.212	0.409	9,244,928
	3	-0.596	15.794	-3.389	184.364	0.316	4,320,463
	5	-0.583	17.734	-8.747	229.126	0.295	6,747,110
	10	-1.342	62.924	-2.748	182.182	0.341	4,215,222

The table reports the mean (Mean), standard deviation (Std), skewness (Skew), and kurtosis (Kurt) of European government bond returns maturing in m years. Figures of mean and standard deviation are multiplied by 10^3. For comparison, the normal distribution has zero skewness and a kurtosis of three. The analysis includes 3,627 (3,149) observations for all countries (Belgium and Greece). LF denotes the Lilliefors test statistic defined as $max|S(x) - CDF|$ with $S(x)$ the empirical cdf and CDF the cumulative distribution function of a normal distribution with mean and standard deviation from the empirical data. For all countries and maturities, p-values are well below 0.001 and not provided here. The critical value for a rejection of the null hypothesis that "the data is normally distributed" at the 1% level is 0.0175 (0.0188) for all countries (Belgium and Greece). JB denotes Jarque-Bera test statistic defined as $N \cdot (Skew^2/6 + (Kurt-3)^2/24)$ with N the number of observations. For all countries and maturities, p-values are well below 0.001 and, again, not provided here. The critical value for a rejection of the null hypothesis that "the data is normally distributed" at the 1% level is 9.4828 (9.5242) for all countries (Belgium and Greece).

sovereign risk tend to have higher skewness (Spain to Greece). There seems to be no clear pattern for the sign of skewness. Further tests are needed to discover whether skewness is important for bond returns.

Assuming normality implies a kurtosis of three ($Kurt = 3$). By contrast, the empirical distributions of **all** bond returns exhibit excess kurtosis ($Kurt >> 3$). Returns of German five-year bonds ($Kurt = 4.708$) and Portuguese five-year bonds ($Kurt = 110.194$) are the least and most heavy tailed, respectively. To sum up, the excess kurtosis implies that the returns depart from normality in the tails and indicates that the Gaussian distribution is an inappropriate assumption. Goodness-of-fit tests will provide a more detailed picture.

The second part of Table 2.1 reports test statistics of Lilliefors and Jarque-Bera tests of normality. The Lilliefors test statistic is defined as $max|S(x) - CDF|$ with $S(x)$ the empirical cdf and CDF the cumulative distribution function of a normal distribution with mean and standard deviation from the empirical data. The critical value for a rejection of the null hypothesis that "the data is normally distributed" at the 1% significance level is 0.0175 (0.0188) for all countries (Belgium and Greece). The bonds with the lowest excess kurtosis and, therefore, with the best fit are German five-year bonds. Since the test statistic for these is well above the critical value, normality is nevertheless rejected.

The Jarque-Bera test statistic is defined as $N \cdot (Skew^2/6 + (Kurt - 3)^2/24)$ with N equal to the number of observations. The critical value for a rejection of the null hypothesis that "the data is normally distributed" at the 1% significance level is 9.4828 (9.5242) for all countries (Belgium and Greece). Similar to the Lilliefors test, Germany's five-year bond offers the best fit. However, the test statistic is well above the critical value of 9.4828 and the null hypothesis is rejected. Portuguese five-year bonds exhibit the highest kurtosis and the worst Jarque-Bera fit. In short, Lilliefors and Jarque-Bera tests overwhelmingly reject the Gaussian distribution for all countries and all maturities at the 1% significance level. The empirical distributions depart from the normal distribution mainly in the tails, which is due to excess kurtosis of bond returns. Figure 2.1 illustrates the difference between the empirical and normal distribution.

As an example, Figure 2.1 shows the fit of the normal distribution for three bonds, selected to be representative of countries with low (Germany), considerable (Spain), and high (Greece) exposure to sovereign risk. The figure illustrates ten-year bonds since this is the most interesting maturity for investors (cf. Codogno et al. 2003; Bernoth et al. 2004; Gomez-Puig 2009). In the first row of Figure 2.1, the histogram of the data and the probability density function of the normal distribution are plotted. The second row shows the difference in frequency between the empirical and normal distributions (cf. Young and Graff 1995) and illustrates the goodness-of-fit results reported in Table 2.1. Although the normal distribution fits the German bond somewhat better than the Spanish bond, it still does not exactly match the empirical distribution. The normal

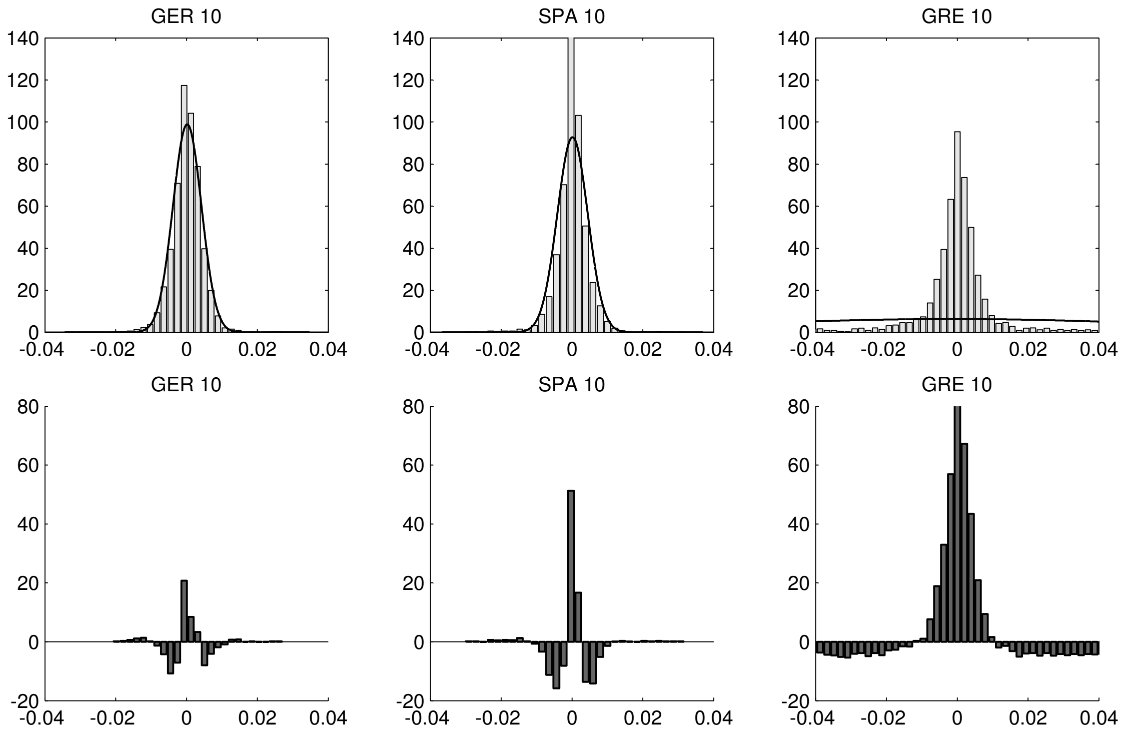

Figure 2.1: Normal distribution and difference in frequency
The figure illustrates the empirical and theoretical distribution of ten year bond returns of Germany, Spain, and Greece in the period of 1999 to 2012: 1. Histograms of daily bond returns and fitted probability density functions of the normal distribution. 2. Difference in frequency between the empirical and normal distribution.

density function underestimates empirical bond returns around the mean and in the tails, while it overestimates them in the shoulders of the distribution. The leptokurtic behavior of Greek bonds prohibits the normal density function from fitting the data in any way.

In sum, the Lilliefors and Jarque-Bera tests overwhelmingly reject the Gaussian distribution for all countries. The graphical analysis in Figure 2.1 shows that the misspecification increases with the country's exposure to sovereign risk. Empirical bond distributions depart from the normal distribution mainly in the tails due to excess kurtosis. The clear rejection of normality forces the investor to consider alternative distributions.

2.3 Alternative Distributions of Bond Returns

The results of the former section lead to the conclusion that the normal distribution is an inappropriate assumption for describing bond returns. However, there is no ex ante financial theory that can aid us in choosing alternative distributions (cf. Aparicio and

Estrada 2001). Therefore, we focus on the empirical distributions' departure from the normal to identify features that the proposed distributions should have. The results of Table 2.1 show that all bonds have considerable excess kurtosis and some bonds exhibit skewness. Therefore, we consider one distribution that exhibits heavy tails and two that account for skewness and heavy tails.

Praetz (1972) and Blattberg and Gonedes (1974) propose Student's *t* distribution for modeling financial returns. The density function of the Student's *t* distribution with unit variance and zero mean is

$$g(X \mid \eta) = \frac{\Gamma\left((\eta+1)/2\right)}{\sqrt{\pi(\eta-2)}\Gamma(\eta/2)}\left(1+\frac{X^2}{\eta-2}\right)^{-(\eta+1)/2} \tag{2.2}$$

with Γ being the Γ-function and $2 < \eta < \infty$ the degrees of freedom. A small value of η implies excess kurtosis. The normal distribution is a special case of Student's *t* distribution if η tends to infinity.

Since Kon (1984), it has been standard practice to account for asymmetry when describing financial returns. Extending the density function in equation (2.2) by a skewness parameter λ results in the skewed Student's *t* distribution, which is able to capture both skewness via λ and excess kurtosis via η. Following Hansen (1994), the density of the skewed Student's *t* distribution is

$$g(X \mid \eta, \lambda) = \begin{cases} bc\left(1+\frac{1}{\eta-2}\left(\frac{bX+a}{1-\lambda}\right)^2\right)^{-(\eta+1)/2}, & X < -a/b, \\ bc\left(1+\frac{1}{\eta-2}\left(\frac{bX+a}{1+\lambda}\right)^2\right)^{-(\eta+1)/2}, & X \geq -a/b \end{cases} \tag{2.3}$$

with $2 < \eta < \infty$ and $-1 < \lambda < 1$. The constants a, b, and c are given by

$$a = \frac{4\lambda c(\eta-2)}{(\eta-1)},$$

$$b^2 = 1 + 3\lambda^2 - a^2, \text{ and}$$

$$c = \frac{\Gamma\left((\eta+1)/2\right)}{\sqrt{\pi(\eta-2)}\Gamma(\eta/2)}.$$

A positive value of λ implies positive skewness and vice versa. By setting $\lambda = 0$, the skewed Student's *t* distribution nests the Student's *t* distribution.[7]

[7]We decide against a more general formulation of the Student's t distribution that allows more extreme values of the tail parameter and the nonexistence of the first moment as using such a formulation would imply that the parameters are no longer comparable to the parameters of the skewed Student's *t* distribution.

The (α-)stable distribution can also exhibit skewness and excess kurtosis (cf. Young and Graff 1995; Rachev and Mittnik 2000). Its four parameters are the index of stability (α), and skewness (β), scale (γ), and location (δ) parameters. In general, there is no closed-form solution, but it is possible to provide the characteristic function. A random variable X is viewed as stable if its characteristic function is (cf. Nolan 2001)

$$E\exp(\mathrm{i}tX) = \begin{cases} \exp\left(-\gamma^{\alpha}|t|^{\alpha}\left[1+\mathrm{i}\beta\left(\tan\frac{\pi\alpha}{2}\right)(\operatorname{sign}t)\left((\gamma|t|)^{1-\alpha}-1\right)\right]+\mathrm{i}\delta t\right), & \alpha \neq 1, \\ \exp\left(-\gamma|t|\left[1+\mathrm{i}\beta\frac{2}{\pi}(\operatorname{sign}t)(\ln|t|+\ln\gamma)\right]+\mathrm{i}\delta t\right), & \alpha = 1 \end{cases} \tag{2.4}$$

with $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, $\gamma > 0$, and $\delta \in \mathbb{R}$. The skewness increases with $|\beta|$. As α tends to 2, the distribution becomes Gaussian and β loses its influence. Lower values of α indicate heavy tails. The second moment does not exist for $\alpha < 2$ or, rather, the variance is infinite. For $\alpha < 1$, the stable distribution has no mean either. In contrast to the skewed Student's *t* distribution, it is not possible to model non-heavy-tailed but skewed returns with this distribution.

The focus of this article is to apply the more popular skewed and fat-tailed distributions, including the Student's *t* and the stable model. These distributions have a long history in academia and are prominent in the financial return literature (see Section 2.1). They provide the foundation for a wide range of commercial applications by leading risk management service providers (cf. Rachev et al. 2010). To ensure consistency, we consider the extension of the Student's *t* distribution by the third moment and include the skewed Student's *t* distribution in our analyses.

2.4 Estimation of Parameters and Goodness-of-Fit Tests

We now present the results of the empirical study. Table 2.2 reports the parameters estimated with maximum likelihood.[8] We assume the returns to be "significantly different from normality" if they:

1. are skewed ($\lambda \neq 0$ for the skewed Student's t or $\beta \neq 0$ for the stable distribution) or

2. exhibit excess kurtosis ($\eta < 30$ for the Student's t and skewed Student's t or $\alpha < 2$ for the stable distribution).

We identify skewness and excess kurtosis if these values are outside the 95% (99%) confidence interval of the estimated parameters, which is indicated by an * (**). The results in Table 2.2 show that the location parameter μ is close to zero and positive for all countries with the exception of Greece. The stable location parameter δ is close to

[8] We first standardize the data for the estimation of the Student's *t* and skewed Student's *t* distribution.

Table 2.2: Parameter estimation for European government bonds

				t	Skewed t		Stable			
	m	μ [10^3]	σ [10^3]	η	λ	η	δ [10^3]	γ [10^3]	β	α
GER	1	0.11	0.41	3.22**	0.05**	3.23**	0.10	0.20	0.13*	1.64**
	3	0.14	1.36	5.22**	-0.02	5.22**	0.17	0.84	-0.12	1.80**
	5	0.17	2.37	6.18**	-0.03	6.35**	0.25	1.51	-0.21*	1.84**
	10	0.23	4.17	5.73**	-0.03	5.88**	0.37	2.63	-0.24**	1.82**
NET	1	0.12	0.64	2.60**	0.01	2.60**	0.11	0.25	0.00	1.43**
	3	0.15	1.56	3.45**	-0.01	3.45**	0.17	0.83	-0.05	1.66**
	5	0.18	2.40	4.36**	-0.04*	4.36**	0.24	1.40	-0.13	1.73**
	10	0.22	6.02	3.53**	-0.03	3.53**	0.35	3.21	-0.10	1.66**
FIN	1	0.12	0.59	2.72**	-0.04**	2.72**	0.14	0.24	-0.09*	1.44**
	3	0.15	1.49	4.79**	-0.04	4.79**	0.19	0.90	-0.13	1.76**
	5	0.18	2.73	4.00**	-0.05*	4.01**	0.28	1.55	-0.19*	1.77**
	10	0.23	5.81	3.43**	-0.04*	3.43**	0.41	3.05	-0.16**	1.67**
FRA	1	0.11	0.53	2.77**	0.06**	2.77**	0.09	0.23	0.12**	1.46**
	3	0.15	1.39	4.30**	-0.01	4.30**	0.17	0.81	-0.07	1.75**
	5	0.17	2.40	4.48**	-0.01	4.48**	0.22	1.42	-0.10	1.79**
	10	0.21	3.98	4.88**	0.00	4.88**	0.23	2.41	0.00	1.81**
AUS	1	0.12	0.66	2.40**	-0.10**	2.37**	0.17	0.17	-0.19**	1.02**
	3	0.15	1.34	2.43**	0.00	2.43**	0.15	0.36	0.00	1.02**
	5	0.18	2.35	2.48**	0.01	2.48**	0.16	0.66	0.00	1.01**
	10	0.23	5.91	2.59**	0.01	2.59**	0.18	1.77	0.00	1.01**
BEL	1	0.12	0.68	2.44**	0.02	2.44**	0.11	0.24	0.07	1.42**
	3	0.17	1.67	3.51**	-0.02	3.51**	0.21	0.89	-0.10	1.69**
	5	0.21	2.64	3.88**	-0.03	3.89**	0.28	1.48	-0.13	1.73**
	10	0.27	4.68	4.30**	-0.03	4.34**	0.40	2.72	-0.14	1.75**
SPA	1	0.12	1.04	2.29**	-0.09**	2.28**	0.19	0.27	-0.13**	1.14**
	3	0.14	2.11	2.29**	-0.04**	2.29**	0.19	0.58	-0.10**	1.21**
	5	0.14	3.12	2.42**	-0.03*	2.42**	0.21	1.02	-0.10**	1.29**
	10	0.12	5.60	2.78**	-0.03*	2.78**	0.27	2.37	-0.08*	1.42**
ITA	1	0.13	0.74	2.32**	-0.01	2.32**	0.13	0.22	0.00	1.33**
	3	0.15	2.15	2.53**	-0.03	2.53**	0.19	0.82	-0.09*	1.49**
	5	0.16	3.41	2.71**	-0.02	2.71**	0.23	1.44	-0.11*	1.56**
	10	0.17	5.61	2.84**	-0.04*	2.84**	0.29	2.50	-0.11*	1.57**
IRE	1	0.15	1.65	2.18**	0.01	2.18**	0.15	0.36	0.00	1.16**
	3	0.18	4.18	2.18**	-0.01	2.18**	0.18	0.93	0.00	1.22**
	5	0.18	4.92	2.32**	-0.02	2.32**	0.26	1.49	-0.08*	1.38**
	10	0.11	11.85	2.24**	-0.04**	2.24**	0.38	3.04	-0.10**	1.26**
POR	1	0.16	2.94	2.10**	0.00	2.10**	0.17	0.38	-0.07**	0.85**
	3	0.17	6.34	2.11**	0.02	2.11**	0.14	0.97	0.00	0.99**
	5	0.15	8.07	2.16**	-0.01	2.16**	0.15	1.60	0.00	1.12**
	10	0.10	17.42	2.24**	-0.01	2.24**	0.12	4.11	0.00	1.10**
GRE	1	-0.63	29.72	2.00**	-0.45**	2.00**	0.12	0.25	0.00	0.77**
	3	-0.60	15.79	2.03**	-0.22**	2.03**	0.21	1.07	-0.11**	0.94**
	5	-0.58	17.73	2.04**	-0.17**	2.04**	0.30	1.67	-0.11**	1.00**
	10	-1.34	62.92	2.01**	-0.16**	2.01**	0.37	3.31	-0.10**	0.99**

The table reports the estimated parameters of the Student's t (t), skewed Student's t (Skewed t), and stable (Stable) distribution. The analysis includes European government bond returns maturing in m years. For the estimation of the Student's t and skewed Student's t distribution, the data is standardized first ((X-μ)/σ). $2 < \eta < \infty$ is the degree of freedom of the Student's t and skewed Student's t distribution. As η tends to infinity the Student's t distribution becomes Gaussian. $-1 < \lambda < 1$ is the skewness parameter of the skewed Student's t distribution. $\delta \in \mathbb{R}$ and $\gamma > 0$ are the location and scale parameters of the stable distribution, respectively. $-1 \leq \beta \leq 1$ and $0 < \alpha \leq 2$ denote its skewness parameter and index of stability. Lower numbers of α indicate heavy tails. As α tends to 2, the distribution becomes Gaussian and β loses its influence. μ, σ, γ, and δ are multiplied by 10^3. An * (**) implies statistical significance of non-normality ($\eta < 30$, $\lambda \neq 0$, $\alpha < 2$, and $\beta \neq 0$) at a 95% (99%) confidence level.

zero and positive for all countries. Remember that the first moment does not exist if $\alpha < 1$, which might be responsible for the difference between the first moment (μ) and the stable location parameter (δ).

Not surprisingly, scale parameters increase with time to maturity and sovereign risk, as do the standard deviations (see Table 2.1). The bond with the lowest (highest) standard deviation is the German one-year (Portuguese ten-year) with $\sigma = 0.41 \cdot 10^{-3}$ ($\sigma = 17.42 \cdot 10^{-3}$). Keeping in mind that the second moment does not exist for the stable distribution, scale parameters still imply the same interactions. They range from $\gamma = 0.20 \cdot 10^{-3}$ for the German one-year bond to $\gamma = 4.11 \cdot 10^{-3}$ for the Portuguese ten-year one.

Skewness parameters of the skewed Student's t distribution are close to zero for all countries with the exception of Greece. Austrian one-year bonds ($\lambda = -0.10^{**}$) have the lowest estimate and French one-year bonds ($\lambda = 0.06^{**}$) the highest. However, the departure from normality ($\lambda \neq 0$) is mainly insignificant. Only 17 (11) of 44 skewness parameters are significantly different from normality at a 95% (99%) confidence level. Estimates of the skewness parameter β of the stable distribution imply similar results. Parameters reflect almost symmetric returns, ranging from $\beta = -0.19^{**}$ (Austrian one-year) to $\beta = 0.12^{**}$ (French one-year). Altogether, only 21 (12) of 44 skewness parameters are significantly different from normality at a 95% (99%) confidence level. In sum, skewness appears to play a minor role in European government bond returns.

All alternative distributions have a kurtosis parameter for providing a better fit in the tails. According to expectations, estimates of the parameter η are almost identical for the Student's t and skewed Student's t distribution (see Section 2.3). Parameter estimates range from $\eta = 6.35^{**}$ (German five-year) to $\eta = 2.10^{**}$ (Portuguese five-year), implying considerably heavy tails. For all countries and maturities, the tail parameter is significantly different from normality at the 99% confidence level. The tail parameters of the stable distribution show similar characteristics. Parameter estimates for α indicate heavy tails among all bond returns and vary between $\alpha = 1.84^{**}$ (German five-year) and $\alpha = 0.85^{**}$ (Portuguese one-year). Analogous to the other distributions, tail parameters are significantly different from normality at a 99% confidence level for **all** countries and maturities, leading to the conclusion that a tail parameter is necessary for matching the characteristics of government bond returns. Goodness-of-fit tests will provide insight into the reliability of the parameter estimation.

Table 2.3 reports statistics and p-values of a χ^2 goodness-of-fit test. The test follows a χ^2 distribution with degrees of freedom depending on the number of parameters and with the null hypothesis that "the empirical distribution equals the distributional assumption." Note that the p-value of the French ten-year bond for the Student's t distribution ($p = 0.355$) is higher than the p-value for the skewed Student's t distribution ($p = 0.336$), although the test statistic is identical ($Stat = 29.64$). Following the argument of parsimony, the difference in p-values is due to the various degrees of freedom.

Table 2.3: χ^2 goodness-of-fit test

	m	t stat	t p	skewed t stat	skewed t p	stable stat	stable p
GER	1	33.11	0.321	26.31	0.554	35.68	0.228
	3	33.51	0.204	32.52	0.198	58.80	0.075
	5	45.44	0.061	41.99	0.076	67.28	0.067
	10	36.79	0.153	32.54	0.235	57.33	0.088
NET	1	47.78	0.147	47.04	0.132	35.31	0.250
	3	33.93	0.295	32.48	0.272	52.13	0.109
	5	24.99	0.614	22.55	0.684	51.32	0.094
	10	35.57	0.191	31.62	0.257	59.95	0.096
FIN	1	89.72	0.002	78.94	0.004	65.43	0.077
	3	51.19	0.035	47.10	0.031	80.13	0.053
	5	72.74	0.004	67.47	0.006	56.03	0.065
	10	42.47	0.101	42.93	0.066	50.08	0.110
FRA	1	71.27	0.011	62.35	0.021	62.47	0.087
	3	23.75	0.701	24.02	0.610	54.42	0.096
	5	31.04	0.326	30.33	0.299	27.01	0.387
	10	29.64	0.355	29.64	0.336	31.71	0.297
AUS	1	221.87	0.001	243.09	0.000	209.07	0.033
	3	385.67	0.000	386.60	0.000	274.61	0.013
	5	444.19	0.000	438.74	0.000	283.37	0.016
	10	422.50	0.000	418.84	0.000	394.94	0.013
BEL	1	68.28	0.044	67.95	0.029	33.11	0.317
	3	22.40	0.760	23.25	0.668	29.34	0.361
	5	31.46	0.349	32.01	0.286	39.35	0.193
	10	34.28	0.210	33.62	0.235	49.19	0.115
SPA	1	242.41	0.002	242.98	0.002	95.35	0.049
	3	157.98	0.007	162.80	0.006	57.98	0.092
	5	109.88	0.007	120.26	0.009	79.56	0.064
	10	129.59	0.001	115.66	0.000	114.19	0.028
ITA	1	108.25	0.020	108.37	0.011	35.58	0.272
	3	48.13	0.156	48.26	0.118	23.69	0.542
	5	51.35	0.066	49.18	0.079	29.64	0.363
	10	61.64	0.014	64.12	0.018	37.25	0.228
IRE	1	241.65	0.011	241.82	0.005	57.06	0.108
	3	178.33	0.006	176.81	0.005	40.27	0.235
	5	115.46	0.014	113.37	0.005	42.71	0.207
	10	206.47	0.004	208.11	0.003	51.12	0.131
POR	1	858.42	0.001	858.40	0.000	147.55	0.030
	3	423.80	0.002	420.14	0.005	70.70	0.096
	5	204.65	0.007	205.04	0.008	42.78	0.201
	10	285.68	0.004	287.42	0.002	74.48	0.090
GRE	1	7,370.08	0.000	38,248.24	0.000	39.30	0.251
	3	1,360.60	0.001	1,571.63	0.000	54.06	0.138
	5	765.19	0.006	913.97	0.002	44.10	0.172
	10	1,772.95	0.001	1,890.03	0.000	31.23	0.265

The table shows the test statistics (Stat) and p-values (p) of a χ^2 goodness-of-fit test of European government bond returns maturing in m years. We test the null hypothesis that "the empirical distribution equals the distributional assumption." The test follows a χ^2 distribution with $p-k-1$ degrees of freedom, where $p = 30$ is the number of intervals and k is the number of parameters estimated for each distribution. Degrees of freedom are 26 for the Student's t (t), 25 the for skewed Student's t (Skewed t) and 25 for the stable (Stable) distribution. For the sake of consistency, all p-values are calculated with simulation techniques for all distributions based on 1,000 repetitions.

The Student's t distribution offers the best fit for the Belgium three-year bond. With a low test statistic of 22.40, the null hypothesis cannot be rejected at a 5% level. The table reports the worst fit for high sovereign risk countries. With a test statistic of 858.42, the null hypothesis has to be rejected for Portuguese one-year bonds. In summary, the Student's t distribution provides a poor fit for European government bond returns. For 24 (17) of a total of 44 bonds, the null hypothesis cannot be rejected at the 1% (5%) significance level.

The skewed Student's t distribution offers similar results. It fits the Belgium three-year bond best and the Portuguese one-year bond worst. Overall, for 22 (17) bonds, the null hypothesis cannot be rejected at the 1% (5%) significance level. This is slightly worse than the fit of the Student's t distribution and somewhat surprising at first sight. Since the skewed Student's t distribution has an additional parameter, one would expect it to provide a better fit. There are two reasons for this underperformance. First, skewness plays only a minor role in European government bonds and the additional parameter does not result in a better fit of the distribution. Second, the poorer p-values might be due to simulated test statistics.

The results of the stable distribution are considerably different. This distribution fits the Italian three-year bond best ($Stat = 23.69$); however, even for the Austria ten-year bond ($Stat = 394.94$), the stable distribution cannot be rejected at the 1% level. In sum, the assumption of stable distributed returns cannot be rejected for any (37) bond(s) at the 1% (5%) level.

Table 2.3 reveals the following conclusions. (1) High exposure to sovereign risk yields a worse fit of the Student's t and skewed Student's t distribution; on the other hand, however, the stable distribution's fit does not seem to depend on sovereign risk. (2) The skewed Student's t provides an even worse fit than the Student's t distribution, meaning that skewness plays a minor role in European government bond returns. (3) The stable distribution outperforms both alternative distributions. At the 1% significance level, it cannot be rejected for any bond.

Figure 2.2 illustrates parameter estimation results and goodness-of-fit tests. Following the design of Figure 2.1, Figure 2.2 presents plots of the goodness of fit of the German, Spanish, and Greek ten-year bonds. Since the stable distribution offers the best fit of the alternative distributions, the figure shows its probability density function and difference in frequency.

In comparison to Figure 2.1, we see a slight improvement in the fitting of the German bond's empirical distribution. The stable distribution captures the peaked returns around the mean better and no longer overestimates the shoulders of the distribution. The plot shows a similarly clear improvement for Spain. A closer look reveals that the stable distribution can capture the departure from normality in the tails. Even for the special case of Greece, the alternative distribution offers a reasonable fit. Considering the difference between the empirical and normal distribution of Greek returns,

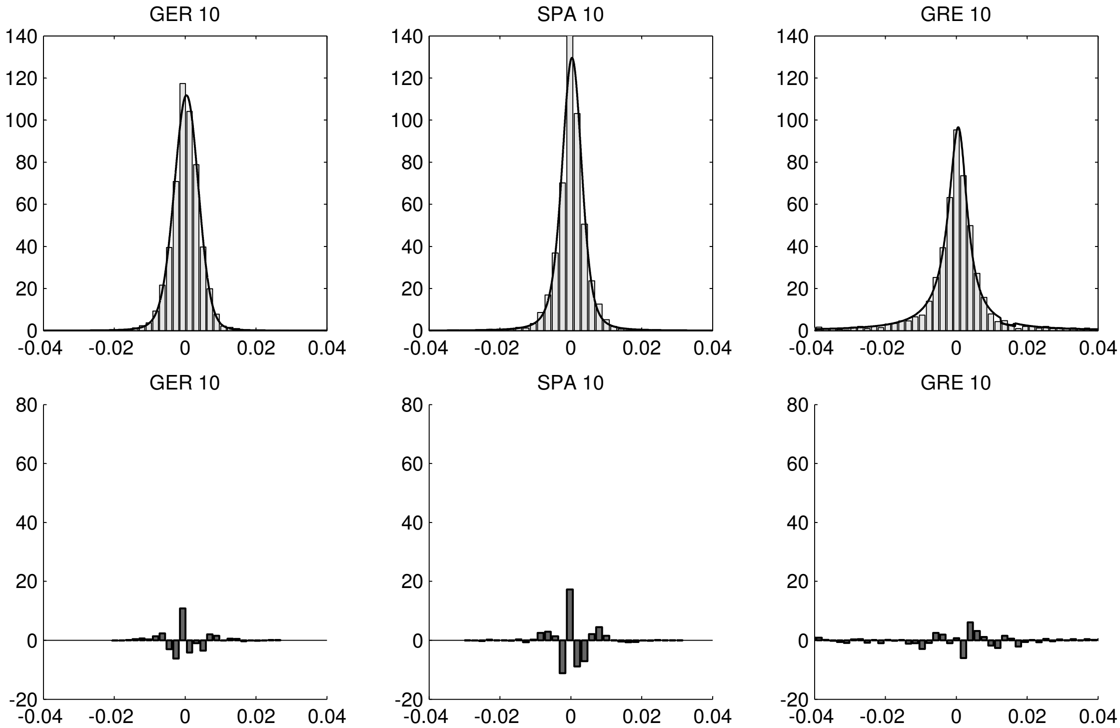

Figure 2.2: Stable distribution and difference in frequency
The figure illustrates the empirical and theoretical distribution of ten year bond returns of Germany, Spain, and Greece in the period of 1999 to 2012: 1. Histograms of daily bond returns and fitted probability density functions of the stable distribution. 2. Difference in frequency between the empirical and stable distribution.

the improvement is remarkable. In particular, the stable distribution does a good job of describing the peaked returns around the mean. Figure 2.2 supports the conclusion drawn from Table 2.3: the stable distribution clearly improves the fit for European bond returns.

2.5 Euro Crisis

From an investor's perspective, the Euro crisis reveals the existence of sovereign risk in Euro bonds. Due to a change of market circumstances, these bond returns begin to behave differently, i.e., the spreads of sovereign risky government bonds increase and yields diverge significantly, leading to the conclusion that structural breaks occur in the time series. The issue to address is whether the significant parameters of skewness and kurtosis are in fact caused by a structural break in the time series. It might well be that (the weak evidence of) skewness is due to a shift in the mean, and that excess kurtosis is the consequence of an increase of standard deviation.

Table 2.4: QLR test results and corresponding headline

	m	date	news
GER NET FIN FRA AUS	1	17/05/2007	Risk and reward: Worried about credit risk? (Flight to quality draws yields to historic low level.)
GER NET FIN FRA	3	02/06/2008	Just bury it: It is time to accept that the Lisbon treaty is dead. The European Union can get along well enough without it. (Lehman bankruptcy rumors)
AUS BEL ITA	3,5 all all	10/11/2011	That's all, folks: For the euro to survive, Italy must not fail. That will require leadership and courage. (Berlusconi resigned)
SPA	all	09/07/2012	The flight from Spain: Spain can be shored up for a while; but its woes contain an alarming lesson for the entire euro zone.
IRE	all	01/07/2011	Can Europe's recovery last? Only if its governments take advantage of sunnier times to make deeper reforms. (Rescue aid for Bank of Ireland approved.)
POR	all	16/01/2012	A false dawn: The recession has been mild so far. But things are likely to get much worse. (S&P downgrade of Portugal)
GRE	1,5	13/10/2011	Nowhere to hide: Investors have had a dreadful time in the recent past. The immediate future looks pretty rotten, too. (Rumors about Greece' debt cut)
	3	04/08/2011	Central bankers to the rescue? They can buy a little time, but the real remedy must come from Western politicians. (Once again the ECB bought government bonds on the secondary market.)
	10	01/12/2011	Is this really the end? Unless Germany and the ECB move quickly, the single currency's collapse is looming. (Greek credit tranche of 8 billion Euro)

The Table reports the date resulting from the QLR test. "all" indicates maturities of one, three, five, and ten years. The news reported are title page headlines of "The Economist". Notes of the authors are given in parenthesis.

Since the existence and, if found, exact date of the structural breaks are unknown, a Quandt likelihood ratio (QLR) statistic is calculated for each time series. To do this, we exclude the first and last three months of the data. For each remaining data point, we perform a Chow (1960) test with null hypothesis being "no structural break in the time series." The maximum of all Chow test statistics corresponds to the QLR test statistic. If the maximum is greater than a critical value at a significance level of 1%, we reject the null hypothesis and report the corresponding date in Table 2.4.[9] Bonds showing a structural break at the same date are grouped in the table. In addition, we report news headlines that might help explain the considerable rise or fall of yields in the right

[9]In this case, we refrain from giving critical values and test statistics.

column of the table. In sum, 31 of 44 time series show evidence of a structural break in the data.

For instance, the table reveals a straightforward link between yield movements and a news headline for Portugal on 14/01/2012. This structural break (Table 2.4; 16/01/2012) is due to a considerable increase in yields in the subsequent period. The increased yields, in turn, might well be caused by the S&P downgrade of Portugal that dominated the newspaper headlines that week.

Subsequently, we reestimate the parameters for the period after the structural break (if one is found). Those periods correspond to the financial crisis in the Eurozone. Like Table 2.2, Table 2.5 reports the maximum likelihood parameter estimates of the Student's t, skewed Student's t, and stable distribution. An * (**) indicates that the parameter is significantly different from normality at a 95% (99%) confidence level.

Depending on the exposure to sovereign risk, Table 2.5 documents a considerable positive shift in location parameters. The mean parameter of Portuguese five-year bonds varies from $\mu = 0.15 \cdot 10^{-3}$ (whole time series) to $\mu = 2.07 \cdot 10^{-3}$ (after structural break). Greece is the only country that experiences a drastic negative shift in the location parameter. Not surprisingly, the standard deviation increases considerably during the period of the Euro crisis and depends on the exposure to sovereign risk. Table 2.5 reports a slight increase of $\sigma = 0.41 \cdot 10^{-3}$ to $\sigma = 0.48 \cdot 10^{-3}$ for German one-year bonds and a clear increase of $\sigma = 17.42 \cdot 10^{-3}$ to $\sigma = 39.77 \cdot 10^{-3}$ for Portuguese ten-year bonds. The same results hold for scale parameters of the stable distribution.

By studying the skewness parameters λ and β, we can discover if a shift of the location parameter in the time series causes the skewness in daily returns. For the slight negative shift of Finnish one-year bonds (from $\mu = 0.12 \cdot 10^{-3}$ to $\mu = 0.10 \cdot 10^{-3}$), the former significant negative skewness parameter ($\lambda = -0.04^{**}$) does indeed become insignificant ($\lambda = -0.03$). In contrast, the skewness parameters of Irish one-, three-, and five-year bonds become significant after the structural break. Altogether, we find weak evidence for skewness in the bonds, which add support to the findings of Section 2.4.

Are the significant excess kurtosis parameters of the whole sample period due to an increase of standard deviation during the Euro crisis period, similar to the findings of the skewness parameters? We find overwhelmingly clear evidence that the answer to this question is no. The kurtosis parameters (λ and α) are still significantly different from normality at a 99% confidence level for **all** time series and **all** alternative distributions. In sum, we find strong evidence for heavy tails even after correcting for structural breaks in the data. In Table 2.6 we report the test statistics and p-values of χ^2 goodness-of-fit tests with null hypothesis being that "the empirical distribution equals the distributional assumption." Since the parameters are estimated for the period after each structural break, the lengths of the time series differ. Hence, the p-values gain credibility because they are invariant to time series length.

Table 2.5: Parameter estimation for European government bonds after structural break

				t	Skewed t		Stable			
	m	μ [10^3]	σ [10^3]	η	λ	η	δ [10^3]	γ [10^3]	β	α
GER	1	0.09	0.48	2.73**	0.11**	2.73**	0.05	0.20	0.23**	1.48**
	3	0.17	1.42	4.68**	0.01	4.69**	0.15	0.85	0.08	1.76**
NET	1	0.10	0.69	2.60**	0.07**	2.60**	0.06	0.27	0.16*	1.44**
	3	0.17	2.04	2.98**	0.01	2.98**	0.14	0.95	0.07	1.54**
FIN	1	0.10	0.72	2.93**	-0.03	2.92**	0.12	0.33	-0.08	1.48**
	3	0.17	1.67	3.83**	0.01	3.83**	0.16	0.93	0.09	1.67**
FRA	1	0.10	0.51	2.86**	0.08**	2.86**	0.06	0.23	0.17*	1.49**
	3	0.18	1.58	3.48**	0.02	3.49**	0.14	0.83	0.10	1.65**
AUS	1	0.10	0.80	2.40**	-0.02	2.40**	0.11	0.23	-0.09*	1.09**
	3	0.17	1.69	2.49**	0.04	2.49**	0.13	0.60	0.11	1.37**
	5	0.32	3.08	2.67**	0.07	2.66**	0.20	1.22	0.20	1.39**
BEL	1	0.11	0.86	2.18**	0.13**	2.18**	0.05	0.19	0.19	1.20**
	3	0.39	2.79	2.32**	0.02	2.33**	0.30	0.81	0.09	1.25**
	5	0.58	4.32	2.53**	0.04	2.53**	0.47	1.57	0.09	1.34**
	10	0.87	7.48	2.70**	-0.02	2.69**	0.88	3.00	0.00	1.37**
ITA	1	0.29	1.83	3.16**	0.06	3.16**	0.22	0.88	0.00	1.48**
	3	0.59	5.40	3.10**	-0.00	3.10**	0.60	2.56	-0.11	1.50**
	5	0.77	8.28	3.27**	0.02	3.28**	0.67	4.13	-0.09	1.56**
	10	1.23	12.46	3.79**	0.02	3.81**	0.77	6.86	0.08	1.65**
IRE	1	0.57	3.41	2.30**	0.11**	2.28**	0.32	0.90	0.19*	1.19**
	3	1.30	9.66	2.25**	0.15**	2.24**	0.44	2.36	0.22*	1.21**
	5	1.52	9.92	2.27**	0.14**	2.25**	0.61	2.30	0.23**	1.07**
	10	1.09	30.00	2.87**	0.02	2.87**	0.95	13.21	0.13	1.45**
POR	1	0.76	5.05	2.54**	0.08	2.52**	0.56	1.82	0.27*	1.35**
	3	1.68	15.03	2.62**	0.11	2.60**	0.59	5.74	0.32*	1.38**
	5	2.07	18.98	3.35**	0.05	3.33**	1.68	9.72	0.12	1.58**
	10	1.36	39.77	3.75**	0.04	3.77**	-0.41	21.64	0.15	1.61**
GRE	1	-6.56	95.60	2.03**	-0.28**	2.02**	-0.26	2.90	-0.09	0.63**
	3	-5.34	45.55	2.36**	-0.13**	2.36**	-1.45	13.70	-0.16	1.29**
	5	-5.31	52.84	2.54**	-0.14**	2.56**	0.48	19.11	-0.15	1.38**
	10	-8.89	193.85	2.11**	-0.15**	2.10**	5.03	26.76	-0.14	0.93**

The table reports the estimated parameters of the Student's t (t), skewed Student's t (Skewed t), and stable (Stable) distribution for the period after the structural break. The analysis includes countries and maturities (m) where the QLR test (see Table 2.4) indicates a structural break, namely Germany, Netherlands, Finland, and France (one and three years), Austria (one, three, and five year), and Belgium, Italy, Ireland, Portugal, and Greece (all maturities). For the estimation of the Student's t and skewed Student's t distribution, the data is standardized first ((X-μ)/σ). $2 < \eta < \infty$ is the degree of freedom of the Student's t and skewed Student's t distribution. As η tends to infinity the Student's t distribution becomes Gaussian. $-1 < \lambda < 1$ is the skewness parameter of the skewed Student's t distribution. $\delta \in \mathbb{R}$ and $\gamma > 0$ are the location and scale parameters of the stable distribution, respectively. $-1 \leq \beta \leq 1$ and $0 < \alpha \leq 2$ denote its skewness parameter and index of stability. Lower numbers of α indicate heavy tails. As α tends to 2, the distribution becomes Gaussian and β loses its influence. μ, σ, γ, and δ are multiplied by 10^3. An * (**) implies statistical significance of non-normality ($\eta < 30$, $\lambda \neq 0$, $\alpha < 2$, and $\beta \neq 0$) at a 95% (99%) confidence level.

Table 2.6: χ^2 goodness-of-fit test after structural break

		t		skewed t		stable	
	m	stat	p	stat	p	stat	p
GER	1	52.05	0.070	33.54	0.314	23.61	0.542
	3	28.20	0.406	28.32	0.372	36.68	0.208
NET	1	58.86	0.046	47.23	0.079	25.76	0.463
	3	33.73	0.257	33.80	0.249	31.93	0.289
FIN	1	79.09	0.006	77.35	0.006	66.95	0.063
	3	58.59	0.009	58.64	0.000	67.33	0.061
FRA	1	61.90	0.021	55.85	0.016	47.42	0.134
	3	23.78	0.674	23.03	0.684	26.60	0.439
AUS	1	127.47	0.005	133.92	0.002	88.61	0.062
	3	48.03	0.131	49.08	0.055	30.07	0.340
	5	43.93	0.131	48.80	0.037	29.50	0.316
BEL	1	73.36	0.058	74.97	0.031	21.69	0.557
	3	41.03	0.268	40.18	0.222	13.63	0.820
	5	42.56	0.196	42.11	0.128	23.24	0.502
	10	34.89	0.317	34.66	0.248	25.69	0.405
ITA	1	47.39	0.047	46.10	0.033	39.84	0.163
	3	35.82	0.192	35.85	0.153	32.58	0.263
	5	42.52	0.084	41.93	0.051	42.09	0.121
	10	32.66	0.248	31.75	0.194	33.47	0.205
IRE	1	64.99	0.072	56.27	0.078	23.09	0.499
	3	64.52	0.082	57.68	0.077	16.16	0.749
	5	129.67	0.009	130.29	0.002	46.94	0.140
	10	47.14	0.073	47.56	0.042	41.05	0.157
POR	1	36.88	0.295	39.28	0.164	17.62	0.693
	3	35.49	0.273	37.55	0.147	20.92	0.609
	5	27.62	0.491	28.73	0.381	27.14	0.328
	10	17.66	0.925	17.39	0.885	18.69	0.692
GRE	1	511.87	0.001	1,303.68	0.000	36.62	0.225
	3	90.00	0.018	66.94	0.020	30.34	0.313
	5	38.65	0.258	26.78	0.582	19.40	0.684
	10	154.40	0.013	172.60	0.004	27.98	0.383

The table shows the test statistics (Stat) and p-values (p) of a χ^2 goodness-of-fit test for countries and maturities (m) where the QLR test (see Table 2.4) indicates a structural break, namely Germany, Netherlands, Finland, and France (one and three year), Austria (one, three, and five year), and Belgium, Italy, Ireland, Portugal, and Greece (all maturities). We test the null hypothesis that "the empirical distribution equals the distributional assumption." The test follows a χ^2 distribution with $p-k-1$ degrees of freedom, where $p = 30$ is the number of intervals and k is the number of parameters estimated for each distribution. Degrees of freedom are 26 for the Student's t (t), 25 the for skewed Student's t (Skewed t) and 25 for the stable (Stable) distribution. For the sake of consistency, all p-values are calculated with simulation techniques for all distributions based on 1,000 repetitions.

Compared to the whole sample period, the goodness-of-fit statistics for the Euro crisis are even more pronounced. The stable distribution, for instance, offers the best fit for Belgian three-year bonds ($p = 0.820$). In comparison to the best fit of the whole series ($p = 0.542$, Italian three-year bonds), this is a clear improvement. The worst fit after the structural break is for Austrian one-year bonds ($p = 0.062$). However, the null hypothesis can no longer be rejected at the 5% level. In contrast, the worst fit of the whole series ($p = 0.013$, Austrian ten-year bonds) is rejected at the 5% level.

Overall, the fit of the alternative distributions clearly improves after correcting for the structural break. The Student's t distribution cannot be rejected at the 1% (5%) significance level for 26 (21) of a total of 31 bonds. We obtain similar results for the skewed Student's t distribution; it cannot be rejected at the 1% (5%) significance level for 25 (19) bonds. The stable distribution cannot be rejected at the 5% significance level for **any** bond. We further conclude that the goodness of fit diminishes with exposure to sovereign risk. After correcting the time series for structural breaks, the mean shifts and some skewness parameters become insignificant. The overall evidence for skewness is weak. The empirical distribution's deviation mainly occurs in the tails. Therefore, excess kurtosis is highly relevant for an alternative distribution: all proposed alternative distributions exhibit this feature. However, the stable distribution clearly offers the best fit.

2.6 Implications for Risk Management

The previous sections underline the importance of considering higher-order moments when describing European government bond returns. We now analyze the consequence for downside risk when assuming different distributions. Since the VaR is the most widely used tool in risk management (cf. Ammann and Reich 2001), we apply VaR calculations when investigating whether the alternative distributions are able to adequately capture the bond risk. Note that the 99% confidence level is crucial for VaR calculations (cf. Berkowitz and O'Brien 2002).

In light of their importance for fixed income management, bonds maturing after ten years are predominantly discussed in finance literature (cf. Codogno et al. 2003; Bernoth et al. 2004; Gomez-Puig 2009). As we do in Figure 2.1 and 2.2, we again examine a country with low (Germany), considerable (Spain), and high (Greece) exposure to sovereign risk. We use a rolling window approach for all three time series to account for the shift in parameters (see Section 2.5). We take the first 100 returns, fit the distribution, and calculate the 0.01 quantile (that is, the VaR at a 99% confidence level for the 101^{st} day). If the return on the 101^{st} day does (not) exceed the VaR, we assign a one (zero) to the date. We continue rolling this window until the end of the time series. Thus, a hitting sequence is generated consisting of ones and zeros.

Three standard techniques dominate analyses of hitting sequences (cf. Christoffersen 2003). First, an unconditional coverage test determines if the expected fraction of VaR violations θ differs significantly from the realized fraction π_{real}. In the limit, the resulting test-statistic

$$LR_{uc} = -2\ln[(1-\theta)^{T_0}\theta^{T_1}/((1-T_1/T)^{T_0}(T_1/T)^{T_1})] \tag{2.5}$$

follows a χ^2 distribution with one degree of freedom. T_1 (T_0) is the number of times the VaR is (not) violated and $T = T_1 + T_0$ is the total number of VaR observations.

The unconditional converge test provides no information on whether VaR violations are clustered. Therefore, we formulate the independence test statistic

$$LR_{ind} = -2\ln[L(\hat{\pi})/((1-\hat{\pi}_{01})^{T_{00}}\hat{\pi}_{01}^{T_{01}}(1-\hat{\pi}_{11})^{T_{10}}\hat{\pi}_{11}^{T_{11}})]. \tag{2.6}$$

$L(\hat{\pi})$ is the likelihood under the alternative hypothesis of equation (2.5), T_{00} (T_{01}) is the number of observations where no (a) VaR violation follows no violation, T_{10} (T_{11}) is the number of observations where no (a) VaR violation follows a violation, $\hat{\pi}_{01} = T_{01}/(T_{00}+T_{01})$, and $\hat{\pi}_{11} = T_{11}/(T_{10}+T_{11})$. As for the unconditional coverage test, the test statistic is χ^2 distributed with one degree of freedom.

The conditional coverage test combines the insights of the unconditional and independence test in one test statistic:

$$LR_{cc} = LR_{uc} + LR_{ind}. \tag{2.7}$$

The test statistic is χ^2 distributed with two degrees of freedom.

Table 2.7 sets out the VaR calculations at the 99% confidence level for ten-year bonds of Germany, Spain, and Greece. π_{real} (θ) is the realized (expected) ratio of VaR violations, p_{unc}, p_{ind}, and p_{cc} represent the p-values of the unconditional, independence, and conditional coverage test with the null hypothesis being that "the VaR model is correct." At the 5% significance level, we reject the skewed Student's t model only for the Spanish bond in case of the unconditional test. This confirms our finding that skewness is not an important issue in European government bond returns. The strong result of the stable distribution, which almost realizes the expected violation rate ($\pi_{real} = 0.0101$ and $\theta = 0.01$), is remarkable. The Student's t and stable distribution cannot be rejected at the 5% significance level for any country and therefore both provide a reliable framework for the VaR calculations. Since the stable distribution offers the best fit and provides the greatest p-values in VaR calculations, we propose using the stable distribution for risk management purposes.

Table 2.7: VaR calculation

	distribution	π_{real}	θ	p_{unc}	p_{ind}	p_{cc}
GER	t	0.0118	0.01	0.2935	0.3273	0.3564
	skewed t	0.0118	0.01	0.2935	0.4999	0.4587
	stable	0.0121	0.01	0.2250	0.3153	0.2893
SPA	t	0.0077	0.01	0.1624	0.1967	0.1638
	skewed t	0.0062	0.01	0.0177	0.6081	0.0528
	stable	0.0101	0.01	0.9655	0.3555	0.6519
GRE	t	0.0135	0.01	0.0753	0.5538	0.1725
	skewed t	0.0121	0.01	0.2756	0.0733	0.1110
	stable	0.0135	0.01	0.0753	0.3022	0.1207

The table reports a comparison of VaR calculations at the 99% confidence level for a maturity of ten years for German, Spanish, and Greek government bond returns assuming the Student's t (t), skewed Student's t (Skewed t), and stable (Stable) distribution. π_{real} (θ) gives the realized (expected) ratio of VaR violations, p_{unc}, p_{ind}, and p_{cc} represent the p-values of the unconditional, independence, and conditional coverage test (cf. Christoffersen 2003) with null hypothesis being that "the VaR model is correct."

2.7 Conclusion

The assumption that financial returns follow a Gaussian distribution, implicitly or explicitly, is frequently made in the finance literature. However, this assumption of normality has important consequences for portfolio theory, derivative pricing, and other financial applications. Whereas a large body of literature is concerned with the empirical distribution of equity returns, little is known about the distribution of bond returns. This is surprising, as international bond markets exceed international equity markets in terms of capitalization.

In the present chapter, we remedy this situation by studying the distribution of daily European government bond returns in the period of 1999–2012. We find that the Lilliefors and Jarque-Bera tests overwhelmingly reject the Gaussian distribution for all countries. The empirical distribution departs from the normal distribution mainly in the tails due to the excess kurtosis of bond returns. Therefore, the kurtosis parameters of the Student's t, skewed Student's t, and stable distribution are highly significant, whereas we find only weak evidence for the significance of skewness parameters.

The goodness-of-fit tests show that sovereign risk is a crucial factor in bond returns. Hence, the importance of flexibility in the tails increases with exposure to sovereign risk. The stable distribution clearly offers the best fit of the tested alternatives. We find a shift in location parameters and a drastic increase of scale parameters caused by the Euro crisis. Due to this shift, some skewness parameters become insignificant and the overall evidence for their influence remains weak. However, even after correcting the time series for structural breaks, excess kurtosis parameters are highly significant. Indeed, the goodness of fit of the stable distribution becomes even better.

Taking excess kurtosis of bond price variations into account has immediate consequences for risk management. We show in a VaR application that risk management clearly improves when assuming the stable distribution for EMU bonds.

3 Measuring Risk in Electricity Forward Returns

This chapter is based on the corresponding working paper by Laitenberger and Lau (2014).

3.1 Introduction

This chapter investigates the risks of an investment in electricity forwards. From a financial point of view, an investment in "electricity" can only be done indirectly by investing in a financial derivative of the underlying asset "electricity." That is, due to its inherent non-storability, the commodity "electricity" is basically useless on the spot market as a financial asset. Nonetheless, investing in electricity forwards could be a portfolio diversification strategy or a means of hedging a given exposure to electricity risk. Since the hedge usually will not be perfect, distribution of the returns of the investment in the electricity forward is of interest. Given the modern approach to risk management, it is especially the risks in the tails of the distribution that are important. These are generally measured with the value at risk (VaR). We study a wide range of different models accounting for time variation in daily returns of one-year and one-quarter forwards. We find that considering ARMA as well as GARCH effects improves VaR calculation and recommend applying a filtered historical simulation using an ARMA(1,0)-GARCH(1,1)-t model.

A forward on electricity is a contract for the delivery of a certain amount of electricity each day of a delivery period, typically one day, one week, one quarter, or one year. Electricity forwards are usually settled in cash against the average spot price of the delivery period, which is why they are sometimes also called *swaps* (cf. Benth and Koekebakker 2008). In markets free of arbitrage, the forward and spot prices of financial assets have a simple analytical relationship. Due to storage costs, this relationship is less straightforward for commodities. However, electricity is non-storable – or, can be stored, but only at an unreasonably high cost – and thus must be consumed the moment it is produced. Consequently, there is no analytical link between forward and spot prices in the case of electricity (cf. Benth and Meyer-Brandis 2009; Pirrong and Jermakyan 1999; Vehviläinen 2002). The absence of seasonal patterns in forward prices in contrast to spot prices (cf. Garcia 2005; Weron 2009) is further proof of this phenomenon. Furthermore, while forward prices generally are non-skewed, Bessembinder and Lemmon (2002) discuss the possibility that large upward spikes in marginal production costs can result in skewed spot prices. Thus, the only sort of relationship that can be established between electricity forwards and spot prices involves the risk premium, that is, the spread between the spot price and the forward price (cf. Benth et al. 2008; Bessembinder and Lemmon 2002; Shawky et al. 2003; Longstaff and Wang 2004).

This chapter describes the distributions of electricity forward returns traded on the NASDAQ OMX Commodities Europe. There is a considerable body of literature on the non-Gaussian behavior of electricity spot prices. For example, Byström (2005) and Chan and Gray (2006) analyze spot prices in an AR-GARCH framework with generalized extreme-value-distributed innovations. Weron (2009) uses regime-switching models to capture jumps and seasonal patterns. In contrast, there is very little literature on forward returns. Solibakke (2010) uses a GARCH(1,1) framework to describe electricity forward returns, but makes no mention as to how the dataset was created. Our work is most closely related to that of Frestad et al. (2010). Frestad et al. (2010) and Andresen et al. (2010) both model the entire forward curve dynamic by fitting a smooth curve approximation to the set of observed real forward prices. This approximated forward curve displays the prices for contracts for the delivery on a single date, while the real forwards are always for delivery over a certain period, with weekly periods common in the short end and yearly contracts common in the long end. In a second step, the authors derive prices for synthetic contracts with delivery over non-overlapping periods and a fixed time to maturity from the curve. Subsequently, they analyze the distributional properties of the time series of these synthetic forwards.

Due to the small number of traded contracts and the limited liquidity of some of them, the results of Frestad et al. (2010) might be affected by their construction of the forward curve. Clearly, there are many ways to construct a forward curve from a finite set of observed forward prices (cf. Fleten and Lemming 2003). Furthermore, the prices of the synthetic forwards will vary depending on the methodology employed for their derivation.[1] The forward curve approach is a good choice when properties of the entire forward curve are described. With respect to risk management, however, a procedure that takes into account the individual characteristics of the forward contracts seems more appropriate. Unlike Frestad et al. (2010), we operate on the real forwards $F(t, T-t)$, where t denotes the time index and T the expiration date of the forward. As t progresses, the time remaining before contract maturation decreases. This is in contrast

[1] Consider, for instance, a small horizontal shift of the curve in Frestad et al.'s Figure 2 (2010, p.62). This would produce a considerable change in the values of the synthetic contracts.

to Frestad et al. (2010), who model synthetic contracts with a fixed time to maturity τ. With real data in discrete time, trend effects like the maturity effect[2] might be present. Hence, we use only real data in our analysis. We compute the returns with successive prices for every forward separately and obtain the returns series a trader would realize when investing in these contracts.

It is well known that financial time series data exhibit time-varying distributional properties, such as volatilities. Therefore, it is natural to apply conditional techniques for estimating the means and volatilities of the returns. We employ a broad set of GARCH processes to model the variance process of forward prices, whereas the mean is described by an ARMA framework. A side benefit of this approach is the elimination of any maturity effects that might be present in the data. Although we first look for models well suited for describing the entire distribution, our main objective is to discover the best model for analyzing the tails of the distribution in terms of a VaR calculation. We use conditional models based on data encompassing a window of 100 days prior to the forecasting period, since the returns display strong shifts in volatility. We find weak autocorrelation in the returns and strong autocorrelation in the squared returns. The best model in terms of VaR calculation is a model that uses an ARMA component to forecast the mean and a GARCH component that takes the persistent volatility into account.

The next section presents the econometric models considered. In Section 3.3, the data are described, along with some essential statistics, after which we demonstrate a simple momentum trading strategy, ending with a comparison of the models. After choosing the best models, we use them in Section 3.4 to calculate and backtest VaR. The last section summarizes our results.

3.2 The Econometric Models

In this section we briefly discuss the different frameworks used to model the dynamics of daily electricity forward returns. In Subsection 3.3.2 we find evidence for autocorrelation in the returns, as well as heavy tails and time-varying volatility. Based on this, we first consider the following standard ARMA(n,m) specification for a series of returns $\{r_t\}$:

$$r_t = \omega + \epsilon_t + \sum_{i=1}^{n} a_i r_{t-i} + \sum_{j=1}^{m} b_j \epsilon_{t-j} \tag{3.1}$$

with $n, m \geq 0$ as the orders of the autoregressive and moving average process, a_i and b_j as their coefficients, and constant ω. ϵ_t is the conditional innovation process with

[2] The maturity effect, sometimes called the Samuelson effect, manifests as an increase in volatility when approaching the end of the contract (cf. Samuelson 1965). The evidence for these effects in commodities is mixed (cf. Bessembinder et al. 1996; Duong and Kalev 2008).

zero mean and variance σ_t^2. In addition to the Gaussian, we also study t-distributed innovations with $\nu > 2$ degrees of freedom. We use three of the most popular models (cf. Engle and Ng 1993; Ederington and Guan 2010) to describe the time-conditional variance σ_t^2:

1) GARCH(p,q) (cf. Bollerslev 1986)

$$\sigma_t^2 = \kappa + \sum_{k=1}^{p} c_k \sigma_{t-k}^2 + \sum_{l=1}^{q} d_l \epsilon_{t-l}^2 \tag{3.2}$$

with a constant $\kappa > 0$. $p, q \geq 1$ are the orders of the GARCH process and $c_k \geq 0, k = 1, .., p$, and $d_l \geq 0, l = 1, .., q$ are their coefficients with $\sum_{k=1}^{p} c_k + \sum_{l=1}^{q} d_l < 1$.

2) EGARCH(p,q) (cf. Nelson 1991)

$$\log \sigma_t^2 = \kappa + \sum_{k=1}^{p} c_k \log \sigma_{t-k}^2 + \sum_{l=1}^{q} d_l \left[\frac{|\epsilon_{t-l}|}{\sigma_{t-l}} - E\left(\frac{|\epsilon_{t-l}|}{\sigma_{t-l}}\right) \right] + \sum_{l=1}^{q} e_l \frac{\epsilon_{t-l}}{\sigma_{t-l}} \tag{3.3}$$

where $E\left(\frac{|\epsilon_{t-l}|}{\sigma_{t-l}}\right) = \begin{cases} \sqrt{2/\pi} & \text{Gaussian innovation distribution,} \\ \sqrt{\frac{\nu-2}{\pi}} \frac{\Gamma(1/2(\nu-1))}{\Gamma(1/2\nu)} & \text{Student innovation distribution,} \end{cases}$ and Γ the Γ function. The EGARCH volatility model allows positive and negative shocks to have different impacts on the conditional volatility. The conditional variance is modeled in a logarithmic way so that no estimation constraints need be imposed to avoid the variance becoming negative.

3) GJR-GARCH(p,q) (cf. Glosten et al. 1993)

$$\sigma_t^2 = \kappa + \sum_{k=1}^{p} c_k \sigma_{t-k}^2 + \sum_{l=1}^{q} d_l \epsilon_{t-l}^2 + \sum_{l=1}^{q} e_l \mathbf{1}_{t-l} \epsilon_{t-l}^2 \tag{3.4}$$

where $\mathbf{1}_{t-l} = \begin{cases} 1 & \epsilon_{t-l} < 0, \\ 0 & \text{otherwise} \end{cases}$, and $d_l + e_l \geq 0, l = 1, .., q$ as well as $\sum_{k=1}^{p} c_k + \sum_{l=1}^{q} d_l + 1/2 \sum_{l=1}^{q} e_l < 1$. The GJR-GARCH is a second way of modeling asymmetries in the variance process. This model deals with the asymmetry of volatility shocks by using different residual signs.

Bollerslev et al. (1992) state that including only lower-order lags in ARMA as well as in GARCH models is sufficient. We follow this advice and apply the three different ARMA-GARCH frameworks with all combinations (with(out) ω, $n, m = 0, .., 3$, and $p, q = 1, 2, 3$) and with Gaussian and t innovations on every time series by maximizing the log likelihood function (LLF). Overall, this results in fitting 1,728 different models for each time series. In addition to the corresponding LLF value, we calculate the Akaike information criterion (AIC) and the Bayesian information criterion (BIC):

$$AIC = -2vLLF + 2NoP \qquad \text{and} \tag{3.5}$$

$$BIC = -2vLLF + NoP * log(NoO) \tag{3.6}$$

where $vLLF$ is the final value of the maximized LLF, NoP the number of parameters estimated in the model, and NoO the number of observations. We prefer the BIC for measuring the goodness of fit of our models because it (i) penalizes free parameters more severely than does the AIC and (ii) takes the NoO into account. Since we want a parsimonious model that has good fitting characteristics for all analyzed time series, we choose the model with the smallest sum of BICs across all 18 time series (one-year contracts) and 28 time series (one-quarter contracts), respectively.

3.3 Empirical Analysis

In this section we present the data and the results of tests for normality and autocorrelation. We then demonstrate a simple momentum trading strategy. Finally, we fit the models to the data.

3.3.1 Data

We analyze the distribution of daily real-world forward returns. The underlying prices are obtained from NASDAQ OMX Commodities Europe. NASDAQ OMX Commodities Europe provides data for one-year, one-quarter, one-week, and one-day forwards. In the case of a one-week forward, for example, the forward is settled in cash against the daily system price on every day of the delivering period, that is, one week. The time series of one-week and, especially, of one-day forwards are rather short because they are only traded for a couple of days before the delivery period starts, yielding too little data for a thorough analysis. Therefore, in this chapter, we concentrate on forward prices for one-year and one-quarter contracts because they provide a sufficiently long trading history – up to five years. We analyze 18 (28) time series of one-year (one-quarter) contracts. The first one-year (one-quarter) contract delivers in the year 1999 (first quarter 2006), the last one-year (one-quarter) contract delivers in the year 2016 (last quarter 2012).[3] In Appendix A.1 a complete list of all forwards including start and end dates of trading as well as the resulting numbers of observations is provided.

[3] In 2006, the predecessor company of NASDAQ OMX Commodities Europe, Nord Pool, changed parts of its product range and in the course of doing so redefined all products. Contracts with a delivery date before 2006 were listed in NOK/Mwh, whereas from 2006 on, prices are given in €/Mwh.

When considering strategies for investing in forward contracts, the strategy's "return" needs to be clearly defined. The forward price is usually set in such a way that only a small margin payment is made at inception of the contract. Taking the return of such a strategy would amount to investing (almost) zero today and receiving a positive or negative payment in the future. Obviously, for such an investment, the return cannot be computed. We thus take the logarithm of the ratio of the forward prices at the two dates, which we call *log returns* in analogy to the standard definition of returns. Using this form of log returns is equivalent to assuming that the investment in the forward position is collateralized with an amount of cash equal to the forward price, which corresponds to a margin of 100%. For the sake of simplicity, we ignore interest earned on the collateral.

3.3.2 Properties of the Data

Table 3.1 shows the statistics of the Lilliefors (1967) test for the one-year forward contracts. On a 1% significance level, the normal hypothesis can be rejected in 17 of 18 cases. The table also shows the results of a Breusch-Godfrey test (cf., e.g., Greene 2003) for autocorrelation in returns and squared returns up to lag three under the null hypothesis of no correlation. Especially when looking at higher lags, there is evidence of autocorrelation in the returns in about half the time series when we assume a significance level of 5%. Thus, an autoregressive framework seems appropriate. An intuitive argument in favor of considering autoregressive effects in the mean is the continuously declining time to maturity (cf. Ng and Pirrong 1994). Assuming, for example, a constant spot price, one would expect a smooth convergence of the forward price toward the spot price in time, as is the case for bonds. Support for this reasoning is found in the empirical work of Shawky et al. (2003), who show that the risk premium between spot and forward prices decreases as time passes. Considering the squared returns, we have to reject the null hypothesis on a 5% significance level for all lags for all time series. This suggests the presence of autocorrelation in the squared returns, which causes volatility clustering. Figure 3.1 illustrates this clustering using the series of ENOYR-07 as an example. Furthermore, the figure reveals particularly high volatility near the end of maturity. This observation contravenes the theory that volatility declines gradually because of the smoothing of expectations. The increase in volatility could be due to the maturity effect, as described in the introduction to this chapter. In summary, autoregressive volatility modeling is the best, if not the only, choice.

Table 3.2 sets out the findings for the one-quarter forward contracts. On a 1% significance level, the normal hypothesis is rejected in 25 of 28 cases. The results of the Breusch-Godfrey test for autocorrelation in the returns are not as straightforward as they were for the one-year contracts. We find evidence for autocorrelation in only about one-third of the time series. In the squared returns, a very high degree of autocorrela-

Table 3.1: Tests for normality and autocorrelation in one-year forward returns

		autocorr returns			autocorr sq returns		
	Lillief.	lag 1	lag 2	lag 3	lag 1	lag 2	lag 3
FWYR-99	*0.053	0.425	0.670	0.817	**0.007	**0.003	**0.004
FWYR-00	**0.088	0.062	*0.049	*0.041	**0.001	**0.000	**0.000
FWYR-01	**0.115	**0.001	**0.000	**0.000	**0.000	**0.000	**0.000
FWYR-02	**0.115	0.191	0.064	*0.024	**0.000	**0.000	**0.000
FWYR-03	**0.173	**0.008	**0.001	**0.000	**0.000	**0.000	**0.000
FWYR-04	**0.095	0.355	*0.038	*0.016	**0.000	**0.000	**0.000
FWYR-05	**0.062	0.502	**0.002	**0.001	*0.017	**0.000	**0.000
ENOYR-06	**0.091	0.233	0.276	0.373	*0.011	*0.026	**0.000
ENOYR-07	**0.102	**0.000	**0.000	**0.001	**0.000	**0.000	**0.000
ENOYR-08	**0.074	**0.010	*0.024	0.051	**0.000	**0.000	**0.000
ENOYR-09	**0.116	0.201	0.433	0.543	**0.000	**0.000	**0.000
ENOYR-10	**0.116	0.784	0.145	0.232	**0.001	**0.000	**0.000
ENOYR-11	**0.089	0.178	**0.000	**0.001	**0.001	**0.000	**0.000
ENOYR-12	**0.069	0.407	*0.028	0.067	*0.011	**0.000	**0.000
ENOYR-13	**0.056	0.792	0.098	0.121	*0.027	**0.001	**0.000
ENOYR-14	**0.073	0.234	0.072	0.152	**0.000	**0.000	**0.000
ENOYR-15	**0.080	0.117	**0.002	**0.002	**0.000	**0.000	**0.000
ENOYR-16	**0.070	0.061	0.180	0.073	**0.000	**0.000	**0.000

The first column of the table shows the name of the product, that is, ENOYR (Electricity Nordic Year) or FWYR (Forward Year), together with the corresponding year (see footnote 3). The second column contains the Lilliefors test statistic on the normal distribution for all one-year forward returns. An * (**) means statistical significance at the 5% (1%) level, meaning that the data are not Gaussian. The other columns show the p-values of a Breusch-Godfrey serial correlation Lagrange multiplier test to detect autocorrelation in the returns and squared returns for lags from one to three under the null hypothesis of no correlation. Again, an * (**) means statistical significance at the 5%(1%) level.

tion appears in all lags for almost all time series, that is, in most instances, the null hypothesis is rejected. The presence of autocorrelation in both cases justifies the use of an autoregressive framework for both, as was the case for one-year contracts.

Table 3.3 contains the average mean, variance, skewness, and kurtosis of the returns of one-year and one-quarter forward contracts across all time series. For the one-year forward contracts, we find mean, variance, and skewness close to zero, but high excess kurtosis. In comparison to the one-year contracts, the one-quarter forward contracts show greater mean, variance, and skewness, but all three moments are still small. The kurtosis is smaller, but still much greater than 3, indicating excess kurtosis. These results are in accordance with those of Frestad et al. (2010) and Shawky et al. (2003), who find non-normality in the forward data in terms of kurtosis, but no asymmetry. Thus, to keep things simple, we refrain from modeling skewness in our analyses.

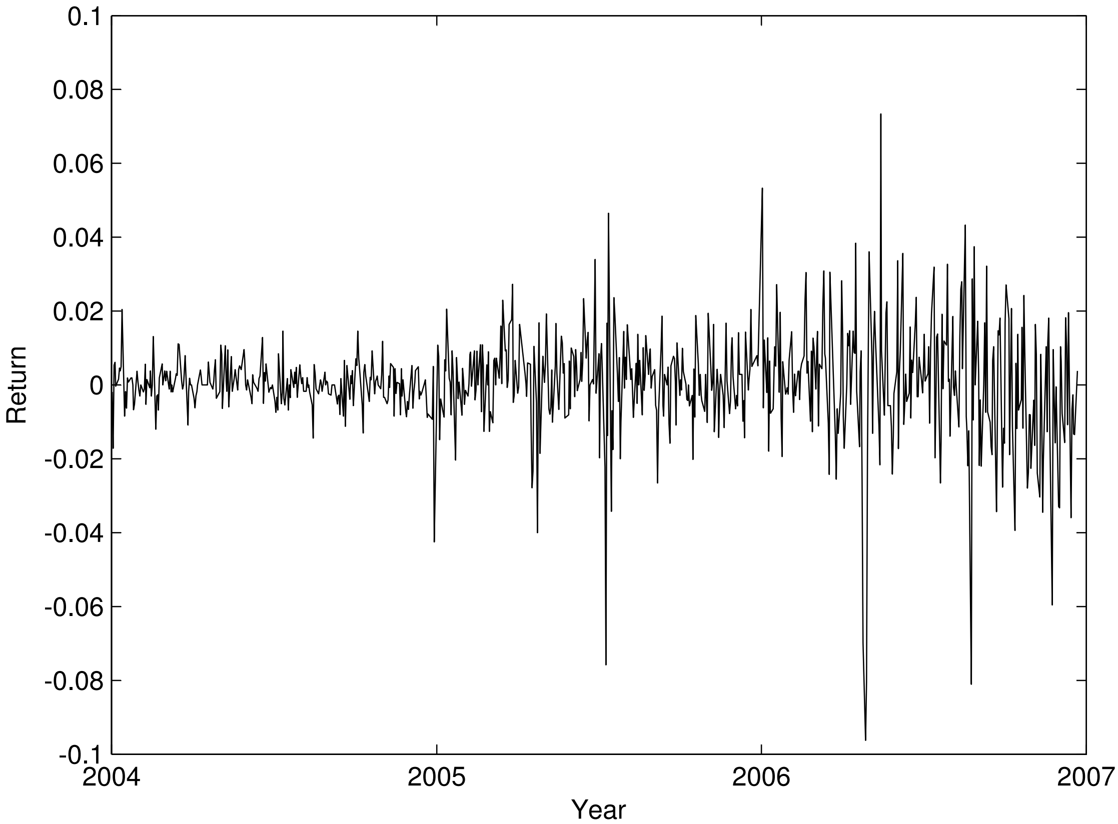

Figure 3.1: Returns of ENOYR-07
The figure shows the log returns of ENOYR-07.

3.3.3 A Simple Momentum Trading Strategy

As a consequence of autocorrelation in the returns, we investigate the possibility of achieving positive returns by applying momentum trading strategies. The finance literature contains a great deal of work on this topic in the context of stock markets. We provide only a brief review of some of the more important contributions. Conrad and Kaul (1988) and Lo and MacKinlay (1989) remark that it is possible to predict short-term returns of stocks. Jegadeesh and Titman (1993) report that momentum investing generates positive returns. Kwon and Kish (2002) use some simple trading rules and outperform a buy-and-hold strategy for the NYSE value-weighted index.

However, the main difference between stock markets and electricity forwards is that stock markets have (or can have) a very long time horizon of several decades while we have only up to five years of electricity trading data. Because we find evidence of autocorrelation in the returns in the last subsection – in other words, we can, to some extent, predict tomorrow's return based on today's return, we try the following

Table 3.2: Tests for normality and autocorrelation in one-quarter forward returns

		autocorr returns			autocorr sq returns		
	Lillief.	lag 1	lag 2	lag 3	lag 1	lag 2	lag 3
ENOQ1-06	**0.107	0.823	0.898	0.799	0.059	**0.001	**0.000
ENOQ2-06	**0.087	**0.001	**0.004	**0.003	**0.000	**0.000	**0.000
ENOQ3-06	**0.088	0.560	0.617	0.809	**0.000	**0.000	**0.000
ENOQ4-06	**0.128	0.072	**0.009	**0.000	**0.000	**0.000	**0.000
ENOQ1-07	**0.127	**0.006	**0.008	**0.006	*0.021	**0.001	**0.000
ENOQ2-07	**0.088	*0.032	0.099	0.122	**0.000	**0.000	**0.000
ENOQ3-07	**0.082	*0.012	*0.036	0.052	**0.000	**0.000	**0.000
ENOQ4-07	**0.060	0.280	0.459	0.615	**0.000	**0.001	**0.000
ENOQ1-08	**0.068	0.239	0.489	0.434	**0.000	**0.000	**0.000
ENOQ2-08	**0.089	0.114	0.101	0.081	**0.000	**0.000	**0.000
ENOQ3-08	**0.088	0.873	0.603	0.328	**0.000	**0.000	**0.000
ENOQ4-08	**0.093	0.457	0.733	0.836	**0.000	**0.000	**0.000
ENOQ1-09	**0.127	0.454	0.601	0.805	**0.000	**0.000	**0.000
ENOQ2-09	**0.100	0.853	0.946	0.255	**0.000	**0.000	**0.000
ENOQ3-09	**0.075	0.456	0.576	0.384	*0.026	**0.000	**0.000
ENOQ4-09	**0.075	0.604	0.867	0.708	**0.008	**0.000	**0.000
ENOQ1-10	**0.053	0.851	0.978	0.971	0.371	*0.014	*0.024
ENOQ2-10	**0.057	0.926	0.772	0.568	0.306	0.120	0.125
ENOQ3-10	**0.061	0.551	*0.029	0.066	0.087	*0.011	**0.003
ENOQ4-10	**0.054	0.976	0.075	0.146	0.066	0.080	*0.026
ENOQ1-11	**0.065	0.071	*0.012	**0.009	0.151	**0.000	**0.000
ENOQ2-11	**0.058	*0.019	*0.028	*0.021	*0.018	**0.000	**0.000
ENOQ3-11	*0.041	0.220	0.472	0.701	0.076	**0.005	**0.001
ENOQ4-11	*0.039	0.506	0.139	0.178	0.118	*0.013	**0.010
ENOQ1-12	**0.047	0.403	0.259	0.299	**0.007	**0.000	**0.001
ENOQ2-12	**0.062	0.179	0.357	0.550	**0.002	**0.001	**0.000
ENOQ3-12	**0.047	0.179	0.218	0.370	**0.004	**0.007	**0.003
ENOQ4-12	*0.035	0.415	**0.010	*0.021	**0.002	**0.004	**0.006

The first column of the table shows the name of the product, that is, ENOQ (Electricity Nordic Quarter), together with the corresponding quarter. The second column contains the Lilliefors test statistic on the normal distribution for all one-quarter forward returns. An * (**) means statistical significance at the 5% (1%) level, meaning that the data are not Gaussian. The other columns show the p-values of a Breusch-Godfrey serial correlation Lagrange multiplier test to detect autocorrelation in the returns and squared returns for lags from one to three under the null hypothesis of no correlation. Again, an * (**) means statistical significance at the 5% (1%) level.

Table 3.3: Mean, variance, skewness, and kurtosis of forward returns

	mean $[10^3]$	variance $[10^3]$	skewness	kurtosis
one year	0.057 $[-0.575; 0.643]$	0.125 $[0.043; 0.237]$	-0.059 $[-0.988; 0.436]$	9.796 $[5.001; 12.577]$
one quarter	0.101 $[-0.598; 0.861]$	0.308 $[0.122; 0.463]$	-0.324 $[-0.778; 0.104]$	6.615 $[4.333; 8.749]$

The table shows the average mean, variance, skewness, and kurtosis of one-year and one-quarter forward returns. Numbers in brackets are 0.15 and 0.85 quantiles. The figures of mean and variance are multiplied by 10^3. Normal-distributed returns would show zero skewness and a kurtosis of 3.

simplest momentum trading strategy. If the first forward return is positive (negative), buy (sell) the forward. If the second forward return is positive (negative) again, stay long (short) in this forward. If the second forward return is negative (positive), close the long (short) position and take a short (long) position in this forward. Continue in this way. We compare this strategy with a simple buy-and-hold strategy. In the case of the one-year forwards, we obtain, in sum, a positive return in 13 of 18 time series with the momentum strategy. The buy-and-hold strategy gives a positive return for only half the time series (9 of 18). The results for the one-quarter forwards are even more noteworthy: with the momentum strategy, 21 of 28 time series yield a positive return, whereas only 13 of 28 time series have a positive return when we use the buy-and-hold strategy. Across all time series, applying a momentum strategy on one-year (one-quarter) forward returns yields an average annual return of 7.6% (11.2%) per time series; a buy-and-hold strategy generates 1.4% (5.1%).

3.3.4 Fitting the Models

In this subsection we discover which models fit the data best. Table 3.4 shows the best 10 models for the one-year forwards in terms of the BIC, as well as the best model with respect to the LLF and the AIC. An ARMA(0,0)-GARCH(1,1)-t specification without ω is the best model for one-year forwards (BIC=-90339.31). This is almost the simplest model and it excludes the ARMA process. Gaussian-based models perform significantly worse than their corresponding t-models; thus, Gaussian models do not appear in the table. Only at the sixth and seventh positions do we begin to find models that

Table 3.4: Models with best goodness of fit for one-year forward returns

model	LLF	AIC	BIC	(rank)
ARMA(0,0)-GARCH(1,1)-t	45406.07	-90668.14	-90339.31	(1)
ARMA(0,0)-GARCH(1,1)-$t(\omega)$	45437.83	-90695.66	-90284.62	(2)
ARMA(0,0)-EGARCH(1,1)-t	45435.17	-90690.34	-90279.30	(3)
ARMA(0,0)-GJR(1,1)-t	45422.58	-90665.16	-90254.12	(4)
ARMA(0,0)-GARCH(2,1)-t	45418.20	-90656.40	-90245.37	(5)
ARMA(0,1)-GARCH(1,1)-t	45416.69	-90653.38	-90242.34	(6)
ARMA(1,0)-GARCH(1,1)-t	45415.92	-90651.84	-90240.80	(7)
ARMA(0,0)-EGARCH(1,1)-$t(\omega)$	45468.80	-90721.61	-90228.36	(8)
ARMA(0,0)-GARCH(1,2)-t	45406.24	-90632.47	-90221.44	(9)
ARMA(0,0)-GJR(1,1)-$t(\omega)$	45454.89	-90693.77	-90200.52	(10)
ARMA(3,3)-EGARCH(3,3)-$t(\omega)$	45701.38	-90754.75	-89275.01	(870)
ARMA(3,1)-EGARCH(3,3)-t	45674.29	-90808.58	-89575.46	(673)

This table shows the best 10 models in terms of the BIC as well as the best AIC and LLF model. In addition to the model specification, the value of the log likelihood function (LLF), the Akaike information criterion (AIC), and the Bayesian information criterion (BIC) are given. All three numbers determine the sum over all one-year forward return time series.

include ARMA parts, which suggests disregarding autoregressive effects in the mean. Incorporating asymmetric volatility in a model generally yields higher LLF values compared to the simpler GARCH model, but the corresponding GARCH model has the smaller BIC value. The last two rows of the table show the best models in terms of LLF value and AIC. In both cases, the ARMA and the GARCH models include more lags. Generally, these models provide a better fit in the LLF value because they are more flexible but, simultaneously, the BIC value always increases. Byström (2003) also uses a GARCH(1,1) model to describe the volatility of forward returns. In contrast to our model, however, he finds that using the Student's t distribution does not yield superior results.

Table 3.5 shows the results for the one-quarter forwards. Again, we find an ARMA(0,0)-GARCH(1,1)-t specification without ω to be the best model on average. Generally, the results are very similar to those for the one-year forwards: incorporating ARMA effects leads to weaker BIC values and applying asymmetric volatility models or including more lags leads to greater LLF values on the one hand, but greater BIC values on the other.

Even though we find evidence of autoregressive effects in the mean (see Subsection 3.3.2), the question is whether an ARMA specification really helps explain the risk of electricity forward returns. To answer this question we conduct a VaR analysis in the next section.

Table 3.5: Models with best goodness of fit for one-quarter forward returns

model	LLF	AIC	BIC	(rank)
ARMA(0,0)-GARCH(1,1)-t	45822.45	-91420.90	-90930.71	(1)
ARMA(0,0)-GARCH(1,1)-$t(\omega)$	45868.37	-91456.75	-90844.01	(2)
ARMA(0,0)-GARCH(2,1)-t	45844.66	-91409.32	-90796.58	(3)
ARMA(0,0)-GJR(1,1)-t	45839.71	-91399.42	-90786.68	(4)
ARMA(0,1)-GARCH(1,1)-t	45835.85	-91391.71	-90778.96	(5)
ARMA(1,0)-GARCH(1,1)-t	45835.53	-91391.05	-90778.31	(6)
ARMA(0,0)-EGARCH(1,1)-t	45834.53	-91389.05	-90776.31	(7)
ARMA(0,0)-GARCH(1,2)-t	45832.36	-91384.73	-90771.99	(8)
ARMA(0,0)-GARCH(2,1)-$t(\omega)$	45891.33	-91446.66	-90711.37	(9)
ARMA(0,0)-GJR(1,1)-$t(\omega)$	45887.90	-91439.80	-90704.51	(10)
ARMA(3,3)-EGARCH(3,3)-$t(\omega)$	46213.76	-91419.51	-89213.65	(1190)
ARMA(2,2)-EGARCH(3,1)-$t(\omega)$	46088.78	-91505.56	-90034.98	(414)

This table shows the best 10 models in terms of the BIC as well as the best AIC and LLF model. In addition to the model specification, the value of the log likelihood function (LLF), the Akaike information criterion (AIC), and the Bayesian information criterion (BIC) are given. All three numbers determine the sum over all one-quarter forward return time series.

3.4 Value at Risk Calculation

In the previous section, we learn that – based on the BIC – we need to refrain from modeling autocorrelation in the returns when taking a parsimonious approach to describe the forward returns ex post; possibly, however, the BIC just overemphasizes the number of parameters in our models. We need to discover whether accounting for autocorrelation in the returns is important for forecasting in an ex ante assessment. Additionally, in this setting, we need to examine which variance model captures the forward returns best. Therefore, we use the 10 best specifications from the last subsection to calculate VaR for confidence levels of 95% and 99% in an out-of-sample study and perform state-of-the-art backtesting. In addition to the basic model, we apply a filtered historical simulation (FHS), as well as extreme value theory (EVT), to the standardized returns. Both models provide a reasonable balance in the tradeoff between feasibility and accuracy and are broadly accepted in the finance literature (cf., e.g., Kuester et al. 2006; Rachev et al. 2010); Solibakke (2010), for example, also applies these models to electricity data.

3.4.1 Value at Risk Methodologies

Calculating the VaR of electricity forward returns is not without some difficulty: (i) in contrast to stocks, for which there are decades' worth of data, time series of electricity forward prices tend to be relatively short; and (ii) with the decreasing time to maturity, a key figure crucial for the valuation changes over time (a stock basically remains the same the whole time). In the following, we briefly discuss three approaches to calculating the VaR.

The first option is to combine different time series to obtain one long series. In regard to point (i) above, this makes sense, but such is not the case in regard to point (ii). Because every time series belongs to a certain product, or rather a certain period of balancing, we would be combining time series of returns that have considerably different maturities.

Let $r_{t,j}$ be the tth return of a time series j. We want to calculate VaR for the return in $t+1$ of j. Another possibility is to use all returns $r_{t+1,1},...,r_{t+1,j-1}$ to compute VaR because, in this case with regard to (ii), we would be looking at the returns of forwards with the same time to maturity. This approach is reasonable if we have a great number of time series j, such as in the case of one-day or one-week forwards. However, this approach is not appropriate for one-year (one-quarter) forwards where we have only 18 (28) time series.

Consequently, although we have shorter time series compared to those for stocks, we calculate VaR with a rolling-window approach by looking at each time series separately and using the following techniques to compute VaR. By fitting the models on the return series, we obtain forecasts for the mean (μ_{t+1}) and standard deviation (σ_{t+1}). Given a confidence level $1-\theta$, this immediately yields the VaR in case of the basic ARMA-GARCH-t model:

$$VaR_{t+1}(\theta) = -\left(\mu_{t+1} + \sigma_{t+1} T_\nu^{-1}(\theta)\right) \tag{3.7}$$

where $T_\nu^{-1}(\theta)$ is the θ quantile of the Student's t distribution with ν degrees of freedom.

The FHS introduced by Hull and White (1998) and Barone-Adesi et al. (1999) combines GARCH-type models with a simulation approach. Barone-Adesi et al. (2002), Giannopoulos and Tunaru (2005), and Pritsker (2006) provide applications for financial data. The returns of the times series are standardized with information from the ARMA-GARCH fitting in order to receive iid returns. Let N be the length of the time series. The standardized empirical $\theta \cdot N$-quantile is denoted as $q_{fhs}(\theta)$. The VaR is calculated as

$$VaR_{t+1}(\theta) = -\left(\mu_{t+1} + \sigma_{t+1} q_{fhs}(\theta)\right). \tag{3.8}$$

The final approach under consideration is based on EVT. EVT has been used on financial data for more than 20 years (cf., e.g., Longin 1996; Gençay and Selçuk 2004; Gilli 2006). The core concept is that only the tails of the distribution are of interest because this is the important region when calculating VaR. Given a threshold u, the distribution of the losses exceeding u[4] can be described with a generalized Pareto distribution $G_{\beta,\xi}(x)$ (for a detailed derivation, see McNeil 1999). This strategy is called *peaks over threshold*. The estimation of $\beta > 0$ and $\xi \geq 0$ is done by MLE. The corresponding quantile $G^{-1}_{\hat{\beta},\hat{\xi}}(\theta)$ is calculated as

$$G^{-1}_{\hat{\beta},\hat{\xi}}(\theta) = u + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{N}{N_u}(1-\theta)\right)^{-\hat{\xi}} - 1\right) \tag{3.9}$$

with N_u as the number of observations exceeding u, $\hat{\beta} > 0$ and $\hat{\xi} \geq 0$. With regard to conditional volatility modeling, McNeil and Frey (2000) present a quasi MLE-based VaR as

$$VaR_{t+1}(\theta) = -\left(\mu_{t+1} + \sigma_{t+1} G^{-1}_{\hat{\beta},\hat{\xi}}(\theta)\right). \tag{3.10}$$

3.4.2 Backtesting

We now describe and apply three state-of-the-art techniques for backtesting VaR (for a detailed description of the following three backtesting methods, see Christoffersen 2003). Taking the first 100 returns of each time series, we fit the models to the data and calculate VaR as described in Subsection 3.4.1. Then, we check whether the return violates the VaR on the next day. If it does (not), we note a 1 (0). Next, we skip the first

[4]We choose u in such a way that 10% of the data exceeds u (cf. Subsection 1.1.4).

and include the 101^{st} return and re-estimate the model. In this way, we apply a rolling window to the whole time series, thus achieving a hitting sequence that is analyzed as follows. First, we perform an unconditional coverage test to determine whether the expected fraction of VaR violations θ differs significantly from the realized fraction π. Under the null hypothesis $\theta = \pi$, the resulting test statistic of the unconditional coverage test is χ^2 distributed with one degree of freedom.

The unconditional coverage test contains no information if VaR violations occur in clusters. For example, if $\theta = 5\%$ and we find five violations in 100 observations we accept a VaR model if we are basing our decision on the unconditional coverage test because, in this case, $\pi = \theta$. If these five violations appear on five days in a row, we will have serious doubts about our model. In the event of finding such a correlation in the violations, we can simply use this information to build a smarter model. To account for this kind of violation clustering, we use the independence test. Under the null hypothesis of independence, the test statistic is χ^2 distributed with one degrees of freedom.

The third test under consideration is the conditional coverage test, which combines both of the above tests. Its test statistic is χ^2 distributed with two degrees of freedom.

This backtesting methodology differs from that of Frestad et al. (2010). Since in their paper a smooth forward curve approximation is used to calculate synthetic forward returns, they refrain from accounting for autocorrelation both in the returns and squared returns. Hence, for them, it is feasible to split the data into two separate, fixed periods, one for estimation and the other for validation of the model. Since we use real-world data, we employ ARMA-GARCH modeling and calculate a conditional forecast for the mean and variance for every single day in a unique way. Thus, we need to employ a different backtesting methodology, one that involves rolling windows.

Table 3.6 shows the evaluation of the VaR hitting sequences of the 10 best models from Table 3.4 on a confidence level of 95% for all one-year forward time series. We also test the best models in terms of LLF and AIC, but they have clearly worse performance and so those results are not shown here. p_{unc}, p_{ind}, and p_{cc} denote the average p-values of the different tests (unconditional coverage, independence, and conditional coverage) over all 18 one-year forward time series and p_{av} gives the average over those three average p-values. To get a picture of the variation, the number in brackets shows for how many of the 18 time series the model is rejected on a 5% significance level. For instance, an average p-value of 0.5 might follow from p-values where one half is 0 and the second half is 1 (this implies rejection of the model in 50% of the cases at any significance level) or results from a series of identical p-values all displaying 0.5. We prefer the latter, of course, because it suggests that the model is feasible for every time series. We also analyze the FHS and EVT approaches described in the last subsection.

In most instances, the basic model performs more poorly than either of the other models. We find that the best model is an ARMA(1,0)-GARCH(1,1)-t in combination

Table 3.6: VaR calculation on a 95% level (one-year forwards)

		p_{unc}	p_{ind}	p_{cc}	p_{av}
ARMA(0,0)-GARCH(1,1)-t	basic	0.232 (8)	0.389 (0)	0.215 (7)	0.279
	FHS	0.624 (1)	0.477 (2)	0.555 (0)	0.552
	EVT	0.627 (1)	0.457 (2)	0.540 (1)	0.542
ARMA(0,0)-GARCH(1,1)-$t(\omega)$	basic	0.224 (8)	0.423 (2)	0.207 (4)	0.284
	FHS	0.572 (1)	0.469 (1)	0.549 (1)	0.530
	EVT	0.574 (1)	0.491 (2)	0.519 (2)	0.528
ARMA(0,0)-EGARCH(1,1)-t	basic	0.364 (4)	0.318 (2)	0.284 (4)	0.322
	FHS	0.112 (10)	0.292 (1)	0.109 (10)	0.171
	EVT	0.101 (9)	0.260 (2)	0.089 (10)	0.150
ARMA(0,0)-GJR(1,1)-t	basic	0.263 (8)	0.353 (3)	0.270 (5)	0.295
	FHS	0.549 (1)	0.528 (2)	0.562 (1)	0.546
	EVT	0.582 (1)	0.456 (1)	0.529 (1)	0.522
ARMA(0,0)-GARCH(2,1)-t	basic	0.222 (10)	0.446 (0)	0.212 (7)	0.293
	FHS	0.567 (1)	0.465 (1)	0.524 (0)	0.519
	EVT	0.576 (1)	0.494 (0)	0.532 (1)	0.534
ARMA(0,1)-GARCH(1,1)-t	basic	0.244 (7)	0.420 (2)	0.291 (7)	0.318
	FHS	0.467 (1)	0.489 (1)	0.501 (1)	0.486
	EVT	0.489 (1)	0.423 (2)	0.459 (2)	0.457
ARMA(1,0)-GARCH(1,1)-t	basic	0.228 (7)	0.388 (1)	0.253 (7)	0.289
	FHS	0.522 (1)	0.635 (0)	0.593 (0)	0.583
	EVT	0.551 (1)	0.496 (2)	0.532 (1)	0.526
ARMA(0,0)-EGARCH(1,1)-$t(\omega)$	basic	0.477 (2)	0.216 (8)	0.227 (7)	0.307
	FHS	0.122 (9)	0.214 (6)	0.105 (13)	0.147
	EVT	0.131 (7)	0.214 (5)	0.109 (11)	0.151
ARMA(0,0)-GARCH(1,2)-t	basic	0.290 (4)	0.434 (1)	0.327 (4)	0.350
	FHS	0.399 (1)	0.475 (0)	0.446 (2)	0.440
	EVT	0.438 (1)	0.422 (1)	0.420 (2)	0.427
ARMA(0,0)-GJR(1,1)-$t(\omega)$	basic	0.372 (2)	0.365 (5)	0.332 (3)	0.356
	FHS	0.541 (1)	0.445 (0)	0.517 (1)	0.501
	EVT	0.554 (1)	0.460 (1)	0.517 (1)	0.510

This table shows the backtesting results of the 10 best models with regard to the whole time series (see Table 3.4) on a confidence level of 95% for all one-year forward time series. p_{uc}, p_{ind}, and p_{cc} denote the average p-values from the different tests (unconditional coverage, independence, and conditional coverage) over all 18 time series and p_{av} gives the average over those three average p-values. The number in brackets is the number of times the model is rejected on a 5% significance level. Backtesting is conducted for the basic model, FHS, and EVT.

with a FHS. This model has the highest p-value for the independence test ($p_{ind} = 0.635$) and the conditional coverage test ($p_{cc} = 0.593$) and returns the greatest average p-value ($p_{av} = 0.583$). Only the p-value of the unconditional coverage test is lower than that of the other models, but it is still high ($p_{unc} = 0.522$). On a 5% significance level, this model has to be rejected only once for the unconditional coverage test and never for the other tests, superior results compared to any other model.

Table 3.7 shows the backtesting results on a confidence level of 99%. Compared to the 95% confidence level, the results are mixed: we find some very strong VaR approaches but also some weak ones, the latter especially when we calculate VaR with the EVT.

Table 3.7: VaR calculation on a 99% level (one-year forwards)

		p_{unc}	p_{ind}	p_{cc}	p_{av}
ARMA(0,0)-GARCH(1,1)-t	basic	0.616 (1)	0.692 (0)	0.669 (0)	0.659
	FHS	0.523 (0)	0.556 (0)	0.564 (0)	0.548
	EVT	0.134 (7)	0.420 (1)	0.182 (5)	0.245
ARMA(0,0)-GARCH(1,1)-$t(\omega)$	basic	0.487 (2)	0.745 (0)	0.623 (0)	0.618
	FHS	0.453 (0)	0.588 (0)	0.544 (0)	0.528
	EVT	0.087 (9)	0.404 (1)	0.140 (4)	0.210
ARMA(0,0)-EGARCH(1,1)-t	basic	0.147 (11)	0.259 (3)	0.113 (11)	0.173
	FHS	0.070 (15)	0.208 (6)	0.064 (15)	0.114
	EVT	0.015 (16)	0.225 (6)	0.026 (16)	0.089
ARMA(0,0)-GJR(1,1)-t	basic	0.499 (1)	0.573 (3)	0.555 (3)	0.542
	FHS	0.419 (1)	0.538 (2)	0.514 (3)	0.490
	EVT	0.058 (13)	0.392 (2)	0.106 (10)	0.185
ARMA(0,0)-GARCH(2,1)-t	basic	0.606 (1)	0.688 (0)	0.691 (0)	0.661
	FHS	0.587 (0)	0.592 (0)	0.651 (0)	0.610
	EVT	0.120 (7)	0.441 (1)	0.200 (5)	0.254
ARMA(0,1)-GARCH(1,1)-t	basic	0.568 (1)	0.702 (0)	0.656 (0)	0.642
	FHS	0.464 (3)	0.562 (0)	0.548 (1)	0.524
	EVT	0.108 (10)	0.390 (1)	0.157 (7)	0.219
ARMA(1,0)-GARCH(1,1)-t	basic	0.599 (1)	0.730 (0)	0.688 (0)	0.672
	FHS	0.422 (4)	0.529 (0)	0.505 (2)	0.485
	EVT	0.113 (9)	0.404 (1)	0.159 (6)	0.225
ARMA(0,0)-EGARCH(1,1)-$t(\omega)$	basic	0.221 (12)	0.390 (2)	0.218 (10)	0.276
	FHS	0.125 (12)	0.224 (8)	0.155 (13)	0.168
	EVT	0.014 (16)	0.198 (5)	0.026 (16)	0.079
ARMA(0,0)-GARCH(1,2)-t	basic	0.573 (0)	0.694 (0)	0.667 (0)	0.645
	FHS	0.456 (4)	0.562 (1)	0.566 (3)	0.528
	EVT	0.098 (10)	0.396 (1)	0.168 (9)	0.220
ARMA(0,0)-GJR(1,1)-$t(\omega)$	basic	0.474 (1)	0.523 (4)	0.522 (2)	0.506
	FHS	0.380 (1)	0.482 (2)	0.433 (3)	0.432
	EVT	0.074 (13)	0.356 (2)	0.110 (12)	0.180

This table shows the backtesting results of the 10 best models with regard to the whole time series (see Table 3.4) on a confidence level of 99% for all one-year forward time series. p_{unc}, p_{ind}, and p_{cc} denote the average p-values from the different tests (unconditional coverage, independence, and conditional coverage) over all 18 time series and p_{av} gives the average over those three average p-values. The number in brackets is the number of times the model is rejected on a 5% significance level. Backtesting is conducted for the basic model, FHS, and EVT.

This could be due to using 100 values to find the 99% VaR, which is not optimum because it means expecting only one VaR violation when using only 10 values for the parametrization. Because it is difficult, in general, for any model to hit exactly one violation in 100 data points, as the 99% confidence level demands, we do not attach much importance to p_{unc}. Again, we favor an ARMA(1,0)-GARCH(1,1)-t framework, but this time it is our basic model. It has the highest average p-value ($p_{av} = 0.672$), highest p-value for the independence test ($p_{ind} = 0.730$), and second highest p-value for the conditional coverage test ($p_{cc} = 0.688$). As stated, we are not overly concerned that p_{unc} is not the best value for this setting, but it is still very high ($p_{unc} = 0.599$).[5]

The results for the one-quarter forward time series of the 10 best models from Table 3.5 can be found in Table A.2 and A.3 in the appendix. To summarize, we again find that an ARMA(1,0)-GARCH(1,1)-t in combination with a FHS is the best model for calculating VaR on a confidence level of 95%. At the 99% level, many models return strong results, among them the same FHS as for the 95% level. Hence, if a risk manager wants to use only one model for all contracts and confidence levels, we recommend the FHS with ARMA(1,0)-GARCH(1,1)-t. Note that this result differs from the result in Subsection 3.3.4 where we fit the whole time series. Although when we fit the whole time series an autoregressive modeling of the mean is negligible, we find that considering these effects improves downside risk management in terms of VaR.

3.5 Conclusion

In this chapter, based on daily data, we calculate the risk of 18 one-year and 28 one-quarter forward return time series from the electricity exchange NASDAQ OMX Commodities Europe. In contrast to other work on this topic, we use real-world data instead of synthetic ones. First, we find excess kurtosis in the data, whereas skewness is negligible. Because of autocorrelation in the returns as well as in the squared returns, we propose using ARMA-GARCH frameworks. When fitting the whole distribution, a simple GARCH(1,1) model with t distributed innovations provides the best average fit in terms of the BIC for both types of forwards. Hence, the question is whether an ARMA modeling of the mean is necessary. Next, we use a rolling-window approach to calculate VaR for confidence levels of 95% and 99% and backtest our computations. We find that considering ARMA effects improves downside risk management in terms of VaR and recommend employing a FHS with an ARMA(1,0)-GARCH(1,1)-t model for both types of forwards.

[5] At this stage, we consider only the downside risk of a forward contract, which is important for the buyer. From the seller's perspective, we need to look at the right side of the distribution because this is the downside-risk-relevant side for sellers. It is straightforward that we find similar results when we calculate the VaR for this side of the distribution because the distribution of returns is symmetric. The results are not given here.

4 A simple NIG-type approach to calculate Value at Risk based on Realized Moments

This chapter is based on the corresponding article by Lau (2015).

4.1 Introduction

Despite well-known drawbacks, such as the lack of subadditivity (cf. Artzner 1999), value at risk (VaR) is still the dominant method in risk management for estimating downside risk. Assuming a certain level of confidence $1-\theta$, $VaR(\theta)_{t+1}$ of an asset is defined as the number such that the probability that the asset's next day return r_{t+1} will be below $-VaR(\theta)_{t+1}$ is θ: $Pr(r_{t+1} < -VaR(\theta)_{t+1}|\Omega_t) = \theta$, where Ω_t is the information set available in t. From a statistical point of view, VaR is closely related to the quantile function F^{-1} of a distribution of returns: $VaR(\theta)_{t+1} = -F^{-1}(\theta)$. Although there are numerous and sophisticated methods for calculating VaR, its computation is often based on simple concepts such as Gaussian assumptions or historical simulations (cf. Pritsker 2006). The inherent complexity of the more accurate models is at least one reason for their infrequent use, with the result that an incorrect (but clear) VaR method is employed. Another, more objective, explanation for avoidance of using the more complex models is model risk, that is, the fewer the parameters employed in an estimate, the less there is to go wrong. Nevertheless, strong empirical evidence from as early as Mandelbrot (1963) argues that the Gaussian distribution is unable to correctly capture the characteristics of financial returns. Gaussian distributions cannot reproduce stylized facts (cf. Cont 2001) of financial time series such as autocorrelation and leptokurtosis, which is particularly problematic when dealing with the distribution tails that typically occur in a VaR.

The end goal of VaR is to determine the regulatory capital required to cover unexpected losses resulting from risks. These regulatory requirements are set out in Basel II/III for banks and in Solvency II for insurance. Because equity is expensive, it is desirable to assign only that amount of it that is necessary (but sufficient) to cover losses. To calculate this amount, we need, essentially, a model that is accurate; better yet would be one that also keeps the VaR as low as possible. This chapter offers an alternative method for calculating VaR that is simple yet comprehensive, keeps the VaR comparably small, and works even in volatile periods. Based on intraday data from the Deutsche Aktien Index (DAX), we compute forecasts for the empirical moments and use them to parameterize the normal inverse Gaussian (NIG) distribution within a method of moments. Once this is complete, obtaining the VaR is easy.

Recent efforts to calculate VaR utilize parametric models based on the theory of *realized variance* (RV). RV (cf. Andersen and Bollerslev 1998; Andersen et al. 2003; Barndorff-Nielsen and Shephard 2002) is defined as the sum of squared intraday returns over a specified time interval, for example, one day.[1] RV is a reliable model-free proxy for the actual (non-observable) variance. To calculate VaR within a given financial return time series, a standard procedure is to include RV as an external variable in an ARCH process so as to model the variance more accurately. This concept, known as GARCH-X, can be traced back to Engle (2002b). Shephard and Sheppard (2010) state that the squared return term in a GARCH-X model has only marginal explanatory power. The authors thus introduce a high-frequency-based volatility (HEAVY) model that calculates tomorrow's variance only as a function of today's variance and today's realized variance. Furthermore, Hansen et al. (2012) formulate the realized GARCH as a generalization of the GARCH-X. Nevertheless, there is some disagreement in the literature as to whether RV actually improves VaR calculations. Giot and Laurent (2004) compare VaR calculations with ARCH models and RV models and find no improvement in the latter. Clements et al. (2008) argue that models incorporating RV improve VaR calculations in the case of currencies. Beltratti and Morana (2005) investigate long-memory models with high-frequency data and identify superior performance.

A smaller strand of the literature takes a different approach and allows higher moments to have a time-varying property. Time-varying moments are not a new concept (cf. Turtle et al. 1994; Bond and Patel 2003), but Brooks et al. (2005) are the first to build a structured framework for a time-varying kurtosis and provide applications for different daily financial returns. Incorporating realized skewness and kurtosis in models is a straightforward expansion of RV, according to Amaya et al. (2013), who use both to predict stock returns. As for time-varying moments, Liu (2012) applies the Cornish and Fisher (1937) (CF) approximation to standardized realized higher moments when calculating VaR. In this chapter, we use the flexible NIG distribution (cf. Barndorff-Nielsen 1997), which provides clearly better results in terms of backtesting.

In Section 4.2 of this chapter, we derive the theoretical background and build the model. Section 4.3 provides a description and analysis of the data. We then compute,

[1]The term *realized volatility*, which is more common in the literature, is merely the positive square root of the realized variance.

backtest, and compare the VaR to the results of other VaR methodologies. The concluding section discusses possible extensions of the presented method.

4.2 Building the Model

4.2.1 Standardized Realized Moments

The availability of intraday data has resulted in this information being integrated in several recent models. Along with meeting the statistical demand of using all available data, RV can also serve as a model-free proxy for the actual (non-observable) variance. $p_{t,i}$ is the ith price on day t. The ith intraday log return on day t is

$$r_{t,i} = \ln p_{t,i} - \ln p_{t,i-1}. \tag{4.1}$$

The realized variance is simply calculated as the sum over N squared returns:

$$RV_t = \sum_{i=1}^{N} r_{t,i}^2. \tag{4.2}$$

With regard to the ordinary moments, generalizing this concept to realized moments yields:

$$RM(or)_t = \sum_{i=1}^{N} r_{t,i}^{or} \tag{4.3}$$

with $RM(or)$ the realized moment of order $or = 1, 2, 3, \ldots$. Obviously, choosing $or = 2$ results in RV_t.

In Subsection 4.2.3, we calculate the parameters of a distribution with a method of moments procedure, for which we need standardized rather than unstandardized moments. The term *standardized realized moments* refers to realized variance, realized skewness, and realized kurtosis, the latter two of which are standardized by definition. Amaya et al. (2013) define standardized realized skewness and kurtosis as follows:

$$RS_t = \frac{\sqrt{N} \sum_{i=1}^{N} r_{t,i}^3}{RV_t^{3/2}} \tag{4.4}$$

and

$$RK_t = \frac{N \sum_{i=1}^{N} r_{t,i}^4}{RV_t^2}. \tag{4.5}$$

The purpose of dividing the realized 3rd (4th) moment by $RV_t^{3/2}$ (RV_t^2) is to achieve standardization similar to that of ordinary moments. Scaling by $\sqrt{N}$ (N) ensures that RS_t (RK_t) is on a daily level.

4.2.2 Forecasting realized moments

In Subsection 4.2.1, we calculate only today's realized moments. To compute the VaR for the next day, we need accurate forecasts of tomorrow's realized moments. Ever since the seminal work of Engle (1982) and Engle and Bollerslev (1986), it has become standard procedure in financial time series analysis to use a time-varying variance. The idea of constructing a time-varying framework for the kurtosis to allow for separate behavior dates back to Brooks et al. (2005). A simple but nonetheless feasible updating scheme is the exponentially weighted moving average (EWMA) adapted by J.P. Morgan and Reuters (1996) to the variance of the daily return r_t:

$$\sigma_{t+1}^2 = \lambda_t \sigma_t^2 + (1 - \lambda_t) r_t^2. \tag{4.6}$$

Given yesterday's variance forecast for today σ_t^2 and today's squared return r_t^2, the variance forecast for tomorrow is σ_{t+1}^2. $0 < \lambda_t < 1$ is the decay factor. To begin with, we use RV_t instead of r_t^2 in Equation (4.6) because RV_t is a better proxy for the recent variance than is the squared return. Applying and extending the EWMA for the remaining higher realized moments is straightforward. According to Liu (2012),

$$M(or)_{t+1} = \lambda_{or,t} M(or)_t + (1 - \lambda_{or,t}) RM(or)_t \tag{4.7}$$

with $M(or)_t$ ($M(or)_{t+1}$) as yesterday's (today's) forecast for the realized moment of order or for today (tomorrow) and $RM(or)_t$ as today's realized moment of order or, although we are primarily interested in $or = 2, 3, 4$. As for $\lambda_{or,t}$, we assess two different strategies. (1) $\lambda_{or,t}$ is fixed to $\lambda = 0.94$ for **all** ors and ts as suggested by RiskMetrics for the variance. If this strategy is undertaken, note that λ is not estimated as in other empirical work and the amount of data needed for the calculation decreases dramatically. (2) Find the optimal $\lambda_{or,t}$s for each equation or in every t by minimizing average squared errors using the information set available up to and including time t.[2]

After forecasts for the realized moments are obtained, we use them to compute forecasts for the standardized realized moments by simply replacing the sums ($\sum_{i=1}^{N} r_{t,i}^{or}$) in Equations (4.2), (4.4), and (4.5) with the forecasts for the realized moments ($M(or)_{t+1}$). We denote those forecasts as v_{t+1} (variance), s_{t+1} (skewness), and k_{t+1} (kurtosis). Finally, we apply them to parameterize a distribution that accounts for higher-order moments. The NIG distribution is described in the next subsection.

[2]To guarantee at least some persistence in the variance process, $\lambda_{or,t}$ can be bounded above 0.5. However, in this study we use $0 < \lambda_{or,t} < 1$.

4.2.3 Normal Inverse Gaussian Distribution

For a distribution to be appropriate for our purposes, it must fulfill the following criteria:

1. have broad acceptance in the finance literature,
2. exhibit higher moments, and
3. provide a simple, closed method of moments estimation.

The NIG distribution meets all three requirements. Because of its flexibility, the NIG distribution is often used in a wide range of financial applications. Forsberg and Bollerslev (2002) formulate a GARCH process with NIG innovations to model daily Euro/U.S. dollar exchange rates. Venter and de Jongh (2002) compare VaR based on the NIG distribution with extreme value theory-VaR and find in favor of the former. Chen et al. (2005) calculate VaR with the NIG distribution for exchange rate and German bank portfolio data using adaptive volatility estimation and show a perfect fit for their model. Chen and Lu (2012) investigate simulated and real-world data and find that NIG-based VaR estimation is robust and accurate for a forecasting horizon of one day. (For other applications, see, e.g., Lillestøl 2000; Aas et al. 2006; Eriksson et al. 2009.) Using a NIG distribution instead of a Gaussian distribution results in a more reasonable modeling of financial returns and therefore contributes to a more realistic VaR calculation.

The NIG distribution (cf., e.g., Paolella 2007) has four parameters: steepness (α), asymmetry (β), scale (δ) and location (μ). The density can be written as

$$f_{NIG}(x;\alpha,\beta,\delta,\mu) = \frac{\alpha\delta\exp\left(\delta\sqrt{\alpha^2-\beta^2}+\beta(x-\mu)\right)}{\pi\sqrt{\delta^2+(x-\mu)^2}}K_1(\alpha\sqrt{\delta^2+(x-\mu)^2}) \tag{4.8}$$

with $x \in \mathbb{R}$, $\alpha > 0$, $0 \leq |\beta| < \alpha$, $\delta > 0$, $\mu \in \mathbb{R}$, and K_1 as the modified Bessel function of the third type with index one. A smaller value of α implies a fat-tailed density. An increasing value of $|\beta|$ yields skewness. Given the parameters of the NIG distribution, it is possible to calculate the central moments as follows (cf. Kalemanova et al. 2007)

$$m = \mu + \delta\frac{\beta}{\gamma}, \tag{4.9}$$

$$v = \delta\frac{\alpha^2}{\gamma^3}, \tag{4.10}$$

$$s = 3\frac{\beta}{\alpha\sqrt{\delta\gamma}}, \text{ and} \tag{4.11}$$

$$k = 3 + 3\left(1 + 4\left(\frac{\beta}{\alpha}\right)^2\right)\frac{1}{\delta\gamma}. \tag{4.12}$$

Parameter estimation of the NIG distribution is typically performed with maximum likelihood, but this is not always feasible because of the complexity of the likelihood (cf. Karlis 2002). By solving Equations (4.9)–(4.12) for the parameters, we obtain a closed analytical method of moments solution under rather fair conditions[3] ($k - \frac{5}{3}s^2 - 3 > 0$ and $3k - 4s^2 - 9 > 0$)[4]:

$$\hat{\mu} = m - \frac{3s\sqrt{v}}{3k - 4s^2 - 9}, \tag{4.13}$$

$$\hat{\delta} = \frac{3^{3/2}\sqrt{v(k - \frac{5}{3}s^2 - 3)}}{3k - 4s^2 - 9}, \tag{4.14}$$

$$\hat{\beta} = \frac{s}{\sqrt{v}(k - \frac{5}{3}s^2 - 3)}, \text{ and} \tag{4.15}$$

$$\hat{\alpha} = \frac{\sqrt{3k - 4s^2 - 9}}{\sqrt{v}(k - \frac{5}{3}s^2 - 3)}. \tag{4.16}$$

When considering short time horizons such as daily, the mean of returns is dominated by the variance. Therefore, rejection of a zero mean return hypothesis is not possible (cf. Christoffersen 2003). Thus, the mean m is set to 0.

4.2.4 Value at Risk

Once we have the forecast for the standardized realized moments (v_{t+1}, s_{t+1}, and k_{t+1}), we parameterize the NIG distribution with Equations (4.13)–(4.16). In this way, we estimate the future return distribution for t+1. On a confidence level $(1 - \theta)$, the VaR for the next day is

$$VaR_{NIG}(\theta)_{t+1} = -F_{NIG}^{-1}(\theta; \hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\mu}) \tag{4.17}$$

where $F_{NIG}^{-1}(\theta; \hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\mu})$ is the θ-quantile of the NIG distribution. This model is called RM-EWMA-NIG. In general, the approach described in this chapter can incorporate any distribution as long as the method of moments can be feasibly implemented.

As in Liu (2012), we can expand the Gaussian distribution to account for higher moments by utilizing the CF approximation (cf. Cornish and Fisher 1937). In this case, the unknown quantile $\tilde{z}_\theta$ can be approximated as

$$\tilde{z}_\theta \approx z_\theta + \frac{1}{6}(z_\theta^2 - 1)s_{t+1} + \frac{1}{24}(z_\theta^3 - 3z_\theta)(k_{t+1} - 3) - \frac{1}{36}(2z_\theta^3 - 5z_\theta)s_{t+1}^2 \tag{4.18}$$

[3]The conditions are rather fair because they simply demand that returns should not be much skewed while their kurtosis is low. However, later we find that high third moments tend to coincide with high fourth moments (see Figure 4.1).

[4]Ralf Werner offers a free NIG Toolbox that includes this computation. Eriksson et al. (2009) provide a solution for the method of moments that differs slightly from the one used here. This difference is simply a result of using excess kurtosis instead of kurtosis.

with z_θ the θ-quantile of the standard normal distribution; consequently,

$$VaR_{CF}(\theta)_{t+1} = -\tilde{z}_\theta \sqrt{v_{t+1}}. \tag{4.19}$$

This model is referred to as RM-EWMA-CF. Note that neither procedure needs a complex numerical optimization algorithm.

We compare the models with one semi-parametric and two parametric models. For this purpose, we choose models that are easy to estimate but also allow higher moments and a time-varying variance. Based on the GARCH-X of Engle (2002b), Shephard and Sheppard (2010) drop the squared return term because it had only marginal explanatory power. They formulate the HEAVY model as

$$v_{t+1} = \omega + \alpha_H RV_t + \beta_H v_t \tag{4.20}$$

with $\omega, \alpha_H \geq 0$ and $\beta_H \in [0, 1)$. The second part of their model provides a specification for RV_{t+1} to calculate multistep-ahead forecasts, which is not required in our case.[5] We calculate

$$VaR_{HEAVY}(\theta)_{t+1} = -t_{\theta,\nu} \sqrt{v_{t+1}} \tag{4.21}$$

where $t_{\theta,\nu}$ is the θ-quantile of the t-distribution with ν degrees of freedom.

We are interested in discovering whether using intraday data actually improves the VaR. Thus, the fourth model under consideration is a GARCH(1,1) framework with t-distributed residuals and daily data (T-GARCH). We replace v_{t+1} in Equation (4.21) with the GARCH forecast to calculate the VaR. A filtered historical simulation (cf. Hull and White 1998; Barone-Adesi et al. 1999) that includes the same T-GARCH process is the final and most complex alternative (T-FHS). The VaR is calculated as in Equation (4.21), but we use the $N \cdot (1-\theta)^{\text{th}}$ value from the list of the descending ordered standardized returns instead of $t_{\theta,\nu}$. This model produces very good results (cf. Kuester et al. 2006). However, in their analysis, Kuester et al. (2006) assume a skewed t instead of a t-distributed innovation process in the GARCH model. Since it is well established that heavy tails are far more crucial in financial modeling than skewness, we refrain from modeling skewness. We use the data from 250 $(1,000)$ days to compute the RV (daily) models. When we fix λ to 0.94, the data for one day are sufficient for calculating the VaR for the RM-EWMA models.

Using intraday data, by definition, means that there is a great deal of data to be processed. What would happen if we instead use end-of-day data in the RM-EWMA models? That is, we could simply replace the realized moments in Equation (4.7) by the daily end-of-day return to the power of the order of the corresponding moment. However, this approach returns very poor p-values in the evaluation process. Squared

[5]Kevin Sheppard's website offers the MFE Toolbox, which includes the code for the HEAVY model.

returns, for instance, tend to be a very noisy indicator of variance (cf. Andersen et al. 2001) and, consequently, computed forecasts will be not reliable. RV – on the other hand – is a reliable proxy for the actual variance and is, therefore, the better choice.

4.3 Empirical Analysis

4.3.1 Data

This analysis uses intraday data for the DAX from January 2, 2006 through December 30, 2011.[6] Note that this time span includes the financial crisis, which was an extremely volatile period. On a typical working day the first (final) value is fixed at 9 a.m. (5:45 p.m.); no trading occurs on weekends or public holidays. The time series provides one unique value for every second.

To simplify the calculation, we abstain from using an elaborate sampling algorithm (see Zhang et al. 2005). As the sampling interval tends toward zero (due to market microstructure noise), the RV is a biased estimator of variance. We sample every 300^{th} value (this corresponds to sampling every 5 minutes).[7]

The unconditional higher-order moments of the end-of-day data show minimal skewness (0.1006) and considerable kurtosis (8.9781), which is typical for financial time series. This supports the use of sophisticated distributions, such as the NIG distribution, that exhibit higher moments to model the data. To decide whether to choose a conditional or unconditional model, we must determine if there is autocorrelation in the data. Figure 4.1 shows the realized moments for the entire period. All three panels are dominated by fluctuations at the end of 2008. The figure nicely reveals how peaks in the realized variance coincide with peaks in both other realized moments. Autocorrelation is obviously an issue for the realized variance and 4^{th} moment, but no concrete conclusion can be drawn as to whether autocorrelation is present in the realized 3^{rd} moment, thus necessitating statistical tests. An analysis of the ACFs and PACFs reveals that the realized variance exhibits strong autocorrelation, while we find weak (considerable) autocorrelation in the realized 3^{rd} (4^{th}) moment.[8]

Table 4.1 shows the p-values of Ljung-Box and Breusch-Godfrey tests with the null hypothesis of no autocorrelation in the realized mean[9], realized variance, and realized 3^{rd} and 4^{th} moment. For the realized mean, the Breusch-Godfrey test clearly suggests acceptance of the null hypothesis for the first three lags. The p-value of the Ljung-Box test is 0.0476, indicating that the null hypothesis is barely rejected, assuming a signif-

[6] The data are provided by the Karlsruhe Institute of Technology (KIT).

[7] This sampling frequency is in accordance with the literature (cf. Andersen and Bollerslev 1998). However, the results are robust to variations in sampling frequency.

[8] ACFs and PACFs are provided in the appendix.

[9] This is the sum over all sampled returns of one day.

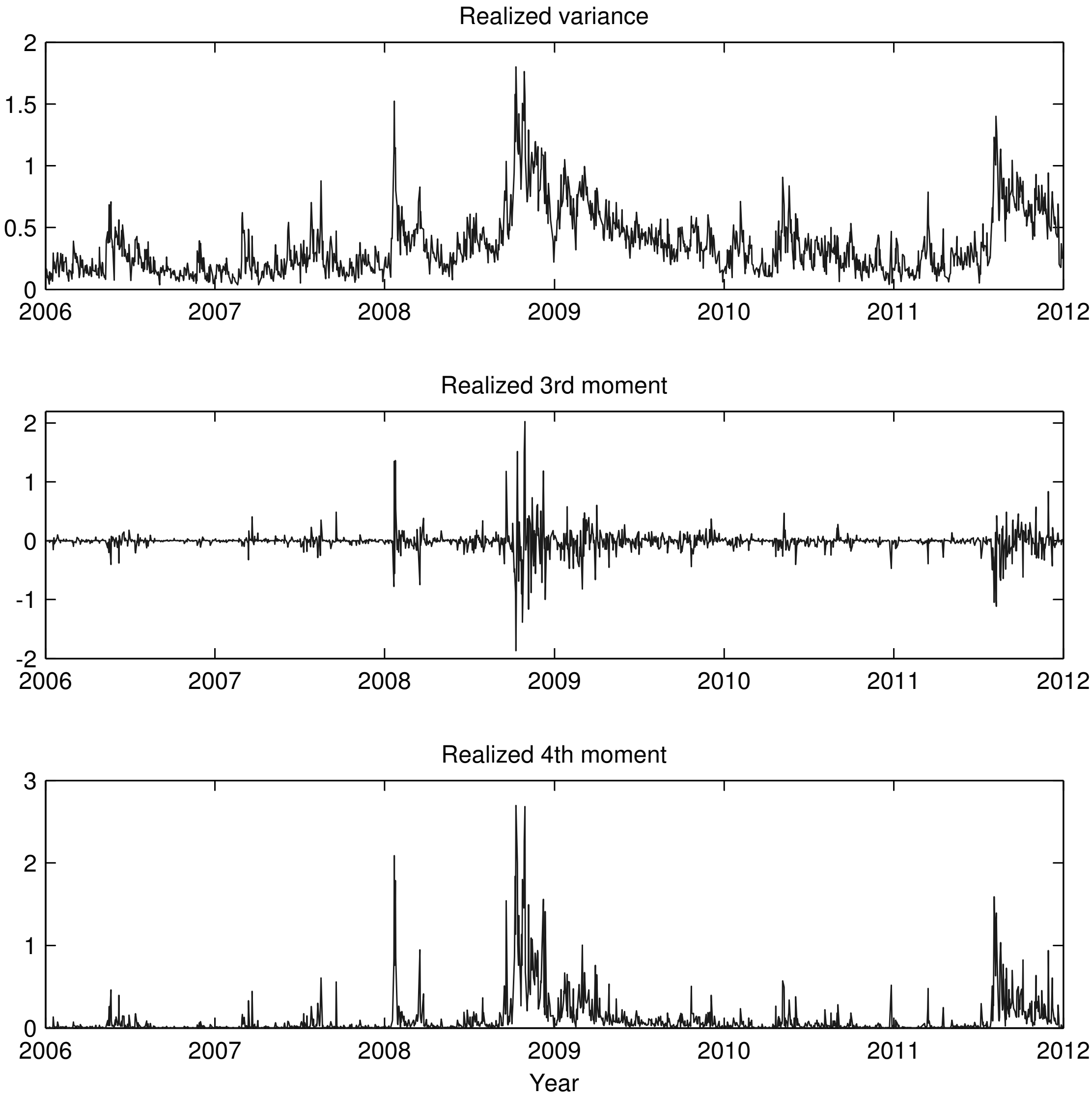

Figure 4.1: Realized moments
The figure shows realized variance, realized 3rd, and realized 4th moment (all on a log scale y-axis for clarity). Corresponding formulas are given in Equation (4.3).

icance level of 0.05. However, due to the clear result from the Breusch-Godfrey test and based on the extant literature, we abandon modeling the realized mean and subsequently assume it to be zero (see Subsection 4.2.3). For all other realized moments, the null hypothesis is rejected at a significance level of 0.05 for all tests. This result differs from Liu (2012), who finds no evidence of autocorrelation in the realized 3rd moment when analyzing the intraday data of IBM. In any event, modeling conditional skewness in this instance comes at no additional cost, in contrast to ARCH processes, where

Table 4.1: Tests on autocorrelation in realized moments

		Breusch-Godfrey		
	LB	lag 1	lag 2	lag 3
Realized mean	0.0476	0.2955	0.4882	0.6997
Realized variance	0.0000	0.0000	0.0000	0.0000
RM(3)	0.0000	0.0050	0.0157	0.0000
RM(4)	0.0000	0.0000	0.0000	0.0000

The table shows p-values of the Ljung-Box (LB) test and Breusch-Godfrey test with lags from 1 to 3 under the null hypothesis of no autocorrelation in the realized moments.

additional parameters always increase the difficulty of estimation. In summary, autocorrelation is a crucial feature of our data. We consider this property by including the EWMA approach (see Subsection 4.2.2) in our model to compute conditional forecasts for the realized moments. In the event the null hypothesis is accepted, we would need to use an unconditional rather than a conditional model.

4.3.2 Backtesting

We calculate the VaR for the next day according to Equations (4.17), (4.19), and (4.21) for all models. If the next day's return violates this VaR, 1 (otherwise 0) is noted. This procedure is carried out for the full time series so as to generate a hitting sequence.

This chapter employs a state-of-the-art technique for backtesting VaR violations. For a detailed description of the following three tests, please refer to Christoffersen (2003). T_0 (T_1) is the number of 0s (1s) and $T = T_0 + T_1$. An unconditional coverage (uc) test is initially performed to determine if the expected fraction of VaR violations θ differs from the realized fraction $\theta_{real} = T_1/T$. Under the null hypothesis $\theta = \theta_{real}$ the likelihood ratio test-statistic

$$LR_{uc} = -2\ln\left[\frac{(1-\theta)^{T_0}\theta^{T_1}}{\left(\frac{T_0}{T}\right)^{T_0}\left(\frac{T_1}{T}\right)^{T_1}}\right] \tag{4.22}$$

is asymptotically χ^2 distributed with one degree of freedom and compared to a critical value of a given significance level. However, the uc test contains no information on a clustering of VaR violations. To check for such dependence in the violations, we conduct the independence (ind) test. Let T_{ij}, $i, j = 0, 1$ be the number of observations where a j follows an i. Under the null hypothesis of independence, the likelihood ratio test statistic is

$$LR_{ind} = -2\ln\left[\frac{\left(\frac{T_0}{T}\right)^{T_0}\left(\frac{T_1}{T}\right)^{T_1}}{\left(\frac{T_{00}}{T_{00}+T_{01}}\right)^{T_{00}}\left(\frac{T_{01}}{T_{00}+T_{01}}\right)^{T_{01}}\left(\frac{T_{10}}{T_{10}+T_{11}}\right)^{T_{10}}\left(\frac{T_{11}}{T_{10}+T_{11}}\right)^{T_{11}}}\right]. \tag{4.23}$$

Again, the statistic is asymptotically χ^2 distributed with one degree of freedom. To account for both demands – correct fraction and independence – simultaneously, the conditional coverage (cc) test combines both of the above tests:

$$LR_{cc} = LR_{uc} + LR_{ind}. \tag{4.24}$$

Consequently, its test statistic is χ^2 distributed with two degrees of freedom.

The fourth backtest is the dynamic quantile (dq) test of Engle and Manganelli (2004)[10]. Given the VaR confidence level $1 - \theta$, we redefine the hitting sequence (see above) by subtracting θ from each item of the sequence. The expected value of this modified hitting sequence, as well as its conditional expectation given any information known at time t, must be zero. Moreover, the modified hitting sequence must be uncorrelated with its lagged values and lagged VaRs. If these conditions are satisfied, the VaR valuations are uncorrelated and will have the correct fraction of violations. To verify this, we set up a regression framework for the modified hitting sequence and include the last four values of both explanatory variables. Under the null hypothesis that all coefficients of this regression are zero, we can then construct a likelihood ratio test that is χ^2 distributed with six degrees of freedom.

4.3.3 Results

Even a simple Gaussian assumption will suffice to calculate VaR for lower confidence levels (cf. Jorion 2007). This is because, even though financial returns are leptokurtic, cdfs of empirical and Gaussian distributions tend to intersect around their 0.05 quantile (this corresponds to the VaR on a 95% confidence level). Therefore, this chapter focuses on higher confidence levels. Table 4.2 shows the results of backtests at confidence levels of $(1 - \theta) = 0.99$, 0.995, and 0.999. p-values are given for uc, ind, cc, and dq tests with the null hypothesis that the VaR model cannot be rejected.

For the 0.99 confidence level, the T-FHS and the RM-EWMA-NIG prove to be the best models. θ_{real} is close to θ, resulting in high p-values for the uc test ($p_{uc} = 0.6101$ and $p_{uc} = 0.5417$). Both models pass the ind, cc, and dq tests with high p-values not less than 0.5. The third best is the HEAVY model, which shows high p-values, except for the dq test ($p_{dq} = 0.2347$). Nevertheless, p_{dq} is still well above a significance level of 0.1. The HEAVY model has the highest p-value for the uc test ($p_{uc} = 0.7334$) because it meets the expected fraction of violations very well ($\theta_{real} = 0.011$). At the significance level of 0.1, we must reject all other models for at least one test. The RM-EWMA-NIG model, where we fix λ, does surprisingly well. Given a 0.05 significance level, we cannot reject the model. The RM-EWMA-CF model fails the dq test ($p_{dq} = 0.0007$). When we fix λ, the model passes the dq test, but still has to be rejected due to weak results in the uc

[10]Simone Manganelli's website offers a code for the dq test.

Table 4.2: Backtests on VaR calculations with confidence levels of 0.99, 0.995, and 0.999

	θ_{real}	θ	p_{uc}	p_{ind}	p_{cc}	p_{dq}
0.99						
T-GARCH	0.0055	0.010	0.0759	0.7812	0.1992	0.7932
T-FHS	0.0086	0.010	0.6101	0.6620	0.7981	0.9695
HEAVY	0.0110	0.010	0.7334	0.5774	0.8080	0.2347
RM-EWMA-CF$_{\lambda=0.94}$	0.0023	0.010	0.0009	0.9054	0.0042	0.1555
RM-EWMA-NIG$_{\lambda=0.94}$	0.0078	0.010	0.4180	0.6911	0.6658	0.0572
RM-EWMA-CF	0.0078	0.010	0.4180	0.6911	0.6658	0.0007
RM-EWMA-NIG	0.0117	0.010	0.5417	0.5504	0.6945	0.9448
0.995						
T-GARCH	0.0023	0.005	0.1339	0.9054	0.3229	0.9233
T-FHS	0.0039	0.005	0.5678	0.8428	0.8329	0.9942
HEAVY	0.0070	0.005	0.3286	0.7208	0.5821	0.8787
RM-EWMA-CF$_{\lambda=0.94}$	0.0016	0.005	0.0418	0.9369	0.1257	0.5936
RM-EWMA-NIG$_{\lambda=0.94}$	0.0039	0.005	0.5678	0.8428	0.8329	0.9531
RM-EWMA-CF	0.0047	0.005	0.8774	0.8119	0.9606	0.0000
RM-EWMA-NIG	0.0039	0.005	0.5678	0.8428	0.8329	0.9977
0.999						
T-GARCH	0.0016	0.001	0.5547	0.9369	0.8373	0.8131
T-FHS	0.0016	0.001	0.5547	0.9369	0.8373	0.6617
HEAVY	0.0016	0.001	0.5547	0.9369	0.8373	0.9186
RM-EWMA-CF$_{\lambda=0.94}$	0.0008	0.001	0.7987	0.9684	0.9672	0.9989
RM-EWMA-NIG$_{\lambda=0.94}$	0.0016	0.001	0.5547	0.9369	0.8373	0.8051
RM-EWMA-CF	0.0063	0.001	0.0001	0.7508	0.0003	0.0000
RM-EWMA-NIG	0.0008	0.001	0.7987	0.9684	0.9672	0.9985

The table shows backtesting results for the T-GARCH, T-FHS, HEAVY, RM-EWMA-CF, and RM-EWMA-NIG models assuming confidence levels of 0.99, 0.995, and 0.999. The subscript $\lambda = 0.94$ indicates that the decay factor is fixed. θ_{real} (θ) represents the realized (expected) fraction of VaR violations. p-values are given for unconditional coverage (uc), independence (ind), conditional coverage (cc), and dynamic quantile (dq) tests with the null hypothesis that the VaR model cannot be rejected: $p \leq 0.01$ (dark gray), $0.01 < p \leq 0.05$ (medium gray), and $0.05 < p \leq 0.1$ (light gray).

and cc tests ($p_{uc} = 0.0009$ and $p_{cc} = 0.0042$). The T-GARCH is only average because it misses the expected fraction of violations ($\theta_{real} = 0.0055$).

The results are not significantly different at the 0.995 confidence level. The T-FHS and the RM-EWMA-NIG models are still the best, but the RM-EWMA-NIG model with fixed λ is not too far behind. All three models pass all tests with high p-values of not less than 0.5. The HEAVY and T-GARCH models also perform better at this level, though the former is still superior. The performance of the RM-EWMA-CF models improves, especially the version with fixed λ, but it is nonetheless still weak. Not surprisingly, because fewer violations are expected and occur, across all models, the p-values of the ind test (and in most instances for the dq test) are high.

At the 0.999 confidence level, all analyzed models do well. Because of hitting θ properly and violations occurring even less frequently, p-values are higher than 0.5 across all models – with one exception. The RM-EWMA-CF model fails three of four tests at a 0.01 significance level because it clearly misses the expected fraction of violations.

In summary, three models cannot be rejected at the 0.1 significance level for all tests: the T-FHS, the RM-EWMA-NIG, and the HEAVY models. The results of the RM-EWMA-NIG model with fixed λ are remarkable, considering that we use no optimization algorithm for the variance.

Figure 4.2 shows the returns and the corresponding VaRs for confidence levels of 0.99, 0.995, and 0.999 for RM-EWMA-NIG and T-FHS, which are the superior models. In general, both models show a similar pattern and adjust rapidly to changes in volatility. The VaR of RM-EWMA-NIG tends to respond more sensitively to massive price fluctuations and appears slightly more "jumpy". This is due to the non-mean-reverting nature of the EWMA, in contrast to the GARCH models. From the perspective of the requirement that a VaR should be as high as necessary but as low as possible, the VaR of RM-EWMA-NIG recovers much faster from shocks, while the VaR of T-FHS tends to remain at a high level. Particularly at the 0.999 confidence levels, both models show a conspicuously different pattern in the last half of 2011, during which there was a long period of high price fluctuation. The VaR of T-FHS during this period looks more jumpy and stays at a high level, whereas the VaR of RM-EWMA-NIG has one high amplitude and decreases more or less constantly to a normal level, which appears to be the more appropriate behavior. In general, the VaR of T-FHS seems to show higher values than the VaR of RM-EWMA-NIG. To quantify this visual impression, we calculate the average VaR as the average over all VaRs. For the important confidence levels of 0.99 and 0.999, the average VaR is much lower for the RM-EWMA-NIG (0.03663 and 0.07107) compared to the T-FHS (0.04008 and 0.09038); that is, utilizing the T-FHS would require too much equity to back up risk compared to the RM-EWMA-NIG.

4.4 Conclusion

Gaussian distributions do not adequately capture the behavior of financial returns, especially when it comes to the tails of a distribution, which is the important region in risk management and for a VaR. The fit is poor because financial returns exhibit kurtosis and occur in volatility clusters. The literature contains many methods designed to take these stylized facts into account, but the models are often complex and sometimes barely feasible. This chapter presents an alternative approach for VaR calculation based on realized moments that even works in extremely volatile periods. We compute forecasts for the realized moments with an EWMA and use them to parameterize the NIG distribution in a method of moments.

Using this technique, we calculate the VaR for the DAX at confidence levels of 0.99, 0.995, and 0.999. Although our alternative VaR methodology is comparatively simple,

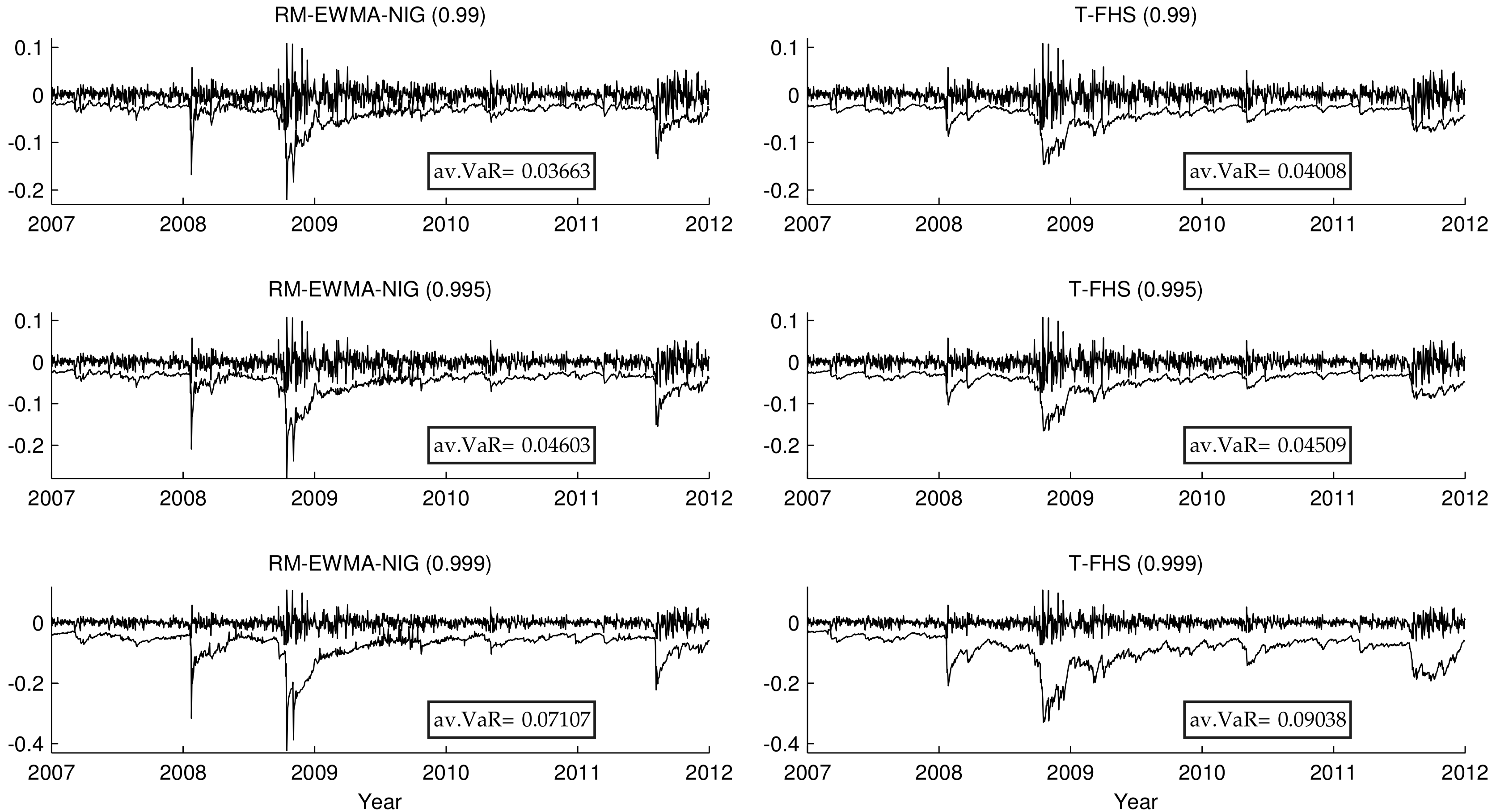

Figure 4.2: Comparison of VaR and returns
The figure shows the calculated VaR based on RM-EWMA-NIG (T-FHS) for confidence levels of 0.99 in left upper (right upper), 0.995 in left middle (right middle), and 0.999 in left lower (right lower) subplots together with the corresponding returns; av.VaR is the average over all VaRs.

our results, in terms of backtesting, stack up well against the long-established T-FHS. Even if we fix the decay factor of the EWMA, and thus simplify the presented method by considerably reducing the amount of data required, the results are still convincing. The VaR is closely linked to the amount of (equity) capital a financial company must provide. In light of the maxim that a VaR should be as high as necessary but as low as possible, the RM-EWMA-NIG model is the best choice for its calculation because its average VaR is lower than the one from the T-FHS.

Although the presented approach already shows very interesting results, there are a few extensions that might further improve it. Replacing the simple EWMA forecast for the moments with a more sophisticated realized GARCH (cf. Hansen et al. 2012), HEAVY, or Corsi (2009) type forecast will allow mean reversion for the moments and enable the derivation of multistep-ahead forecasts. But doing so will, of course, require more complex estimation techniques. Although the NIG distribution is widely used for financial data, we could use other distributions that meet the criteria outlined in Subsection 4.2.3. For example, one possible second candidate is the tempered stable distribution. A convenient feature of the method presented in this chapter is that it can encompass a portfolio point of view with realized co-variance, co-skewness, and co-kurtosis.

A Appendix Chapter 3

Table A.1: Length of one-year and one-quarter forward time series

one-year	Start	End	Obs.	one-quarter	Start	End	Obs.
FWYR-99	10/27/1997	12/28/1998	291	ENOQ1-06	01/02/2004	12/30/2005	502
FWYR-00	10/28/1997	12/28/1999	540	ENOQ2-06	01/02/2004	03/31/2006	567
FWYR-01	09/07/1998	12/27/2000	578	ENOQ3-06	01/02/2004	06/30/2006	625
FWYR-02	03/01/1999	12/21/2001	706	ENOQ4-06	01/02/2004	09/29/2006	690
FWYR-03	01/17/2000	12/23/2002	734	ENOQ1-07	03/01/2005	12/29/2006	502
FWYR-04	01/02/2001	12/23/2003	742	ENOQ2-07	03/01/2005	03/30/2007	566
FWYR-05	01/02/2002	12/28/2004	745	ENOQ3-07	03/01/2005	06/29/2007	625
ENOYR-06	01/02/2003	12/28/2005	748	ENOQ4-07	03/01/2005	09/28/2007	690
ENOYR-07	01/02/2004	12/27/2006	751	ENOQ1-08	01/02/2006	12/28/2007	501
ENOYR-08	01/03/2005	12/21/2007	750	ENOQ2-08	01/02/2006	03/31/2008	562
ENOYR-09	01/02/2006	12/23/2008	751	ENOQ3-08	01/02/2006	06/30/2008	625
ENOYR-10	06/15/2006	12/28/2009	891	ENOQ4-08	01/02/2006	09/30/2008	691
ENOYR-11	06/15/2006	12/28/2010	1,142	ENOQ1-09	01/02/2007	12/30/2008	502
ENOYR-12	01/02/2007	12/28/2011	1,256	ENOQ2-09	01/02/2007	03/31/2009	565
ENOYR-13	01/02/2008	06/06/2012	1,115	ENOQ3-09	01/02/2007	06/30/2009	624
ENOYR-14	01/02/2009	06/06/2012	863	ENOQ4-09	01/02/2007	09/30/2009	690
ENOYR-15	01/04/2010	06/06/2012	612	ENOQ1-10	01/02/2008	12/30/2009	503
ENOYR-16	01/03/2011	06/06/2012	360	ENOQ2-10	01/02/2008	03/31/2010	566
				ENOQ3-10	01/02/2008	06/30/2010	625
				ENOQ4-10	01/02/2008	09/30/2010	691
				ENOQ1-11	01/02/2009	12/30/2010	503
				ENOQ2-11	01/02/2009	03/31/2011	567
				ENOQ3-11	01/02/2009	06/30/2011	626
				ENOQ4-11	01/02/2009	09/30/2011	692
				ENOQ1-12	01/04/2010	12/30/2011	505
				ENOQ2-12	01/04/2010	03/30/2012	570
				ENOQ3-12	01/04/2010	06/06/2012	612
				ENOQ4-12	01/04/2010	06/06/2012	612

This table shows start and end dates of all one-year and one-quarter forward time series. Obs. is number of observations.

Table A.2: VaR calculation on a 95% level (one-quarter forwards)

		p_{unc}	p_{ind}	p_{cc}	p_{av}
ARMA(0,0)-GARCH(1,1)-t	basic	0,390 (6)	0,371 (3)	0,295 (8)	0,352
	FHS	0,579 (0)	0,391 (6)	0,446 (3)	0,472
	EVT	0,625 (0)	0,379 (5)	0,464 (2)	0,489
ARMA(0,0)-GARCH(1,1)-$t(\omega)$	basic	0,455 (6)	0,374 (4)	0,335 (9)	0,388
	FHS	0,566 (0)	0,371 (6)	0,439 (5)	0,459
	EVT	0,568 (0)	0,398 (7)	0,432 (5)	0,466
ARMA(0,0)-GARCH(2,1)-t	basic	0,394 (6)	0,372 (3)	0,300 (8)	0,355
	FHS	0,555 (0)	0,394 (5)	0,460 (4)	0,470
	EVT	0,569 (0)	0,410 (5)	0,469 (2)	0,483
ARMA(0,0)-GJR(1,1)-t	basic	0,448 (6)	0,316 (5)	0,316 (9)	0,360
	FHS	0,465 (2)	0,477 (4)	0,458 (4)	0,467
	EVT	0,471 (0)	0,529 (3)	0,479 (2)	0,493
ARMA(0,1)-GARCH(1,1)-t	basic	0,351 (6)	0,332 (3)	0,259 (8)	0,314
	FHS	0,587 (0)	0,443 (4)	0,515 (3)	0,515
	EVT	0,514 (0)	0,467 (4)	0,503 (4)	0,495
ARMA(1,0)-GARCH(1,1)-t	basic	0,358 (6)	0,358 (3)	0,285 (8)	0,334
	FHS	0,610 (0)	0,432 (4)	0,534 (3)	0,526
	EVT	0,579 (0)	0,437 (4)	0,513 (2)	0,510
ARMA(0,0)-EGARCH(1,1)-t	basic	0,286 (6)	0,273 (8)	0,224 (10)	0,261
	FHS	0,099 (16)	0,314 (7)	0,096 (18)	0,170
	EVT	0,099 (16)	0,334 (4)	0,107 (16)	0,180
ARMA(0,0)-GARCH(1,2)-t	basic	0,366 (7)	0,419 (3)	0,351 (8)	0,379
	FHS	0,539 (2)	0,466 (5)	0,464 (2)	0,490
	EVT	0,502 (1)	0,472 (4)	0,462 (3)	0,478
ARMA(0,0)-GARCH(2,1)-$t(\omega)$	basic	0,475 (6)	0,365 (4)	0,352 (8)	0,397
	FHS	0,518 (0)	0,377 (5)	0,418 (4)	0,438
	EVT	0,506 (0)	0,424 (5)	0,429 (4)	0,453
ARMA(0,0)-GJR(1,1)-$t(\omega)$	basic	0,470 (3)	0,287 (5)	0,315 (6)	0,358
	FHS	0,493 (2)	0,389 (3)	0,421 (4)	0,434
	EVT	0,507 (2)	0,369 (3)	0,414 (4)	0,430

This table shows the backtesting results of the 10 best models with regard to the whole time series (see Table 3.5) on a confidence level of 95% for all one-quarter forward time series. p_{unc}, p_{ind}, and p_{cc} denote the average p-values from the different tests (unconditional coverage, independence, and conditional coverage) over all 28 time series and p_{av} gives the average over those three average p-values. The number in brackets is the number of times the model is rejected on a 5% significance level. Backtesting is conducted for the basic model, FHS, and EVT.

Table A.3: VaR calculation on a 99% level (one-quarter forwards)

		p_{unc}	p_{ind}	p_{cc}	p_{av}
ARMA(0,0)-GARCH(1,1)-t	basic	0,565 (0)	0,666 (2)	0,691 (1)	0,641
	FHS	0,603 (0)	0,614 (3)	0,656 (1)	0,624
	EVT	0,229 (5)	0,440 (2)	0,296 (3)	0,322
ARMA(0,0)-GARCH(1,1)-$t(\omega)$	basic	0,505 (0)	0,653 (2)	0,635 (1)	0,598
	FHS	0,662 (0)	0,630 (3)	0,686 (1)	0,659
	EVT	0,216 (4)	0,464 (1)	0,289 (3)	0,323
ARMA(0,0)-GARCH(2,1)-t	basic	0,609 (0)	0,651 (1)	0,702 (1)	0,654
	FHS	0,592 (0)	0,615 (3)	0,665 (1)	0,624
	EVT	0,217 (5)	0,444 (2)	0,294 (4)	0,318
ARMA(0,0)-GJR(1,1)-t	basic	0,552 (2)	0,539 (6)	0,542 (4)	0,544
	FHS	0,525 (2)	0,577 (3)	0,546 (3)	0,549
	EVT	0,183 (7)	0,455 (2)	0,236 (7)	0,291
ARMA(0,1)-GARCH(1,1)-t	basic	0,521 (0)	0,622 (2)	0,592 (1)	0,578
	FHS	0,573 (0)	0,603 (2)	0,628 (1)	0,601
	EVT	0,236 (8)	0,451 (2)	0,304 (7)	0,331
ARMA(1,0)-GARCH(1,1)-t	basic	0,536 (0)	0,638 (2)	0,610 (1)	0,595
	FHS	0,555 (0)	0,607 (2)	0,633 (1)	0,598
	EVT	0,197 (7)	0,441 (2)	0,278 (7)	0,306
ARMA(0,0)-EGARCH(1,1)-t	basic	0,047 (20)	0,276 (4)	0,058 (20)	0,127
	FHS	0,029 (20)	0,277 (5)	0,037 (21)	0,114
	EVT	0,010 (26)	0,311 (6)	0,017 (26)	0,113
ARMA(0,0)-GARCH(1,2)-t	basic	0,582 (1)	0,654 (2)	0,706 (1)	0,647
	FHS	0,554 (2)	0,623 (2)	0,636 (2)	0,604
	EVT	0,183 (7)	0,420 (1)	0,240 (6)	0,281
ARMA(0,0)-GARCH(2,1)-$t(\omega)$	basic	0,526 (0)	0,649 (1)	0,654 (1)	0,610
	FHS	0,635 (0)	0,628 (2)	0,690 (1)	0,651
	EVT	0,201 (8)	0,464 (1)	0,279 (4)	0,315
ARMA(0,0)-GJR(1,1)-$t(\omega)$	basic	0,445 (3)	0,538 (4)	0,520 (5)	0,501
	FHS	0,543 (2)	0,551 (4)	0,536 (3)	0,543
	EVT	0,185 (7)	0,448 (2)	0,251 (9)	0,295

This table shows the backtesting results of the 10 best models with regard to the whole time series (see Table 3.5) on a confidence level of 99% for all one-quarter forward time series. p_{unc}, p_{ind}, and p_{cc} denote the average p-values from the different tests (unconditional coverage, independence, and conditional coverage) over all 28 time series and p_{av} gives the average over those three average p-values. The number in brackets is the number of times the model is rejected on a 5% significance level. Backtesting is conducted for the basic model, FHS, and EVT.

B Appendix Chapter 4

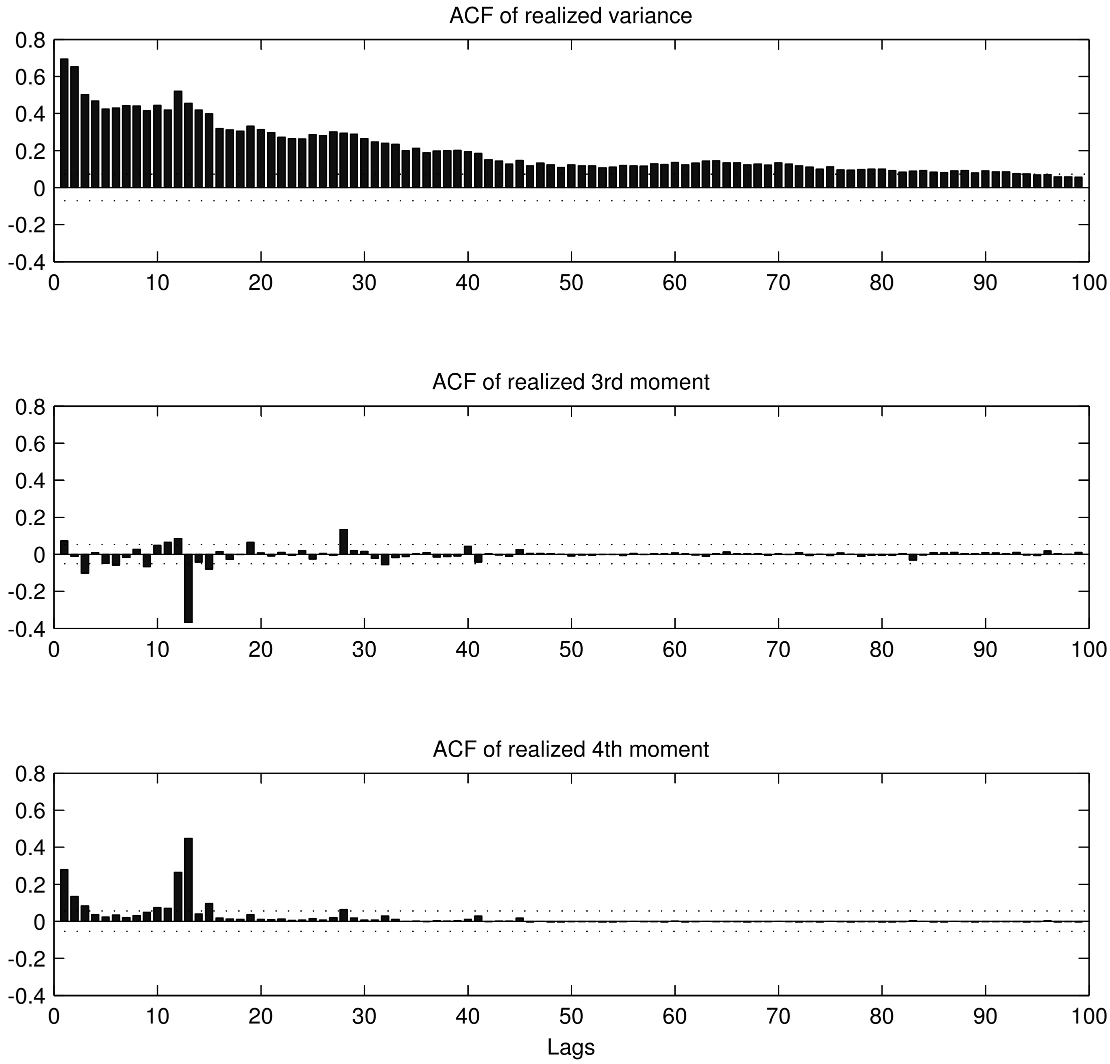

Figure B.1: Autocorrelation functions

The figure shows the autocorrelation function of realized variance, realized 3rd and realized 4th moment. Dotted lines define the 95% confidence interval.

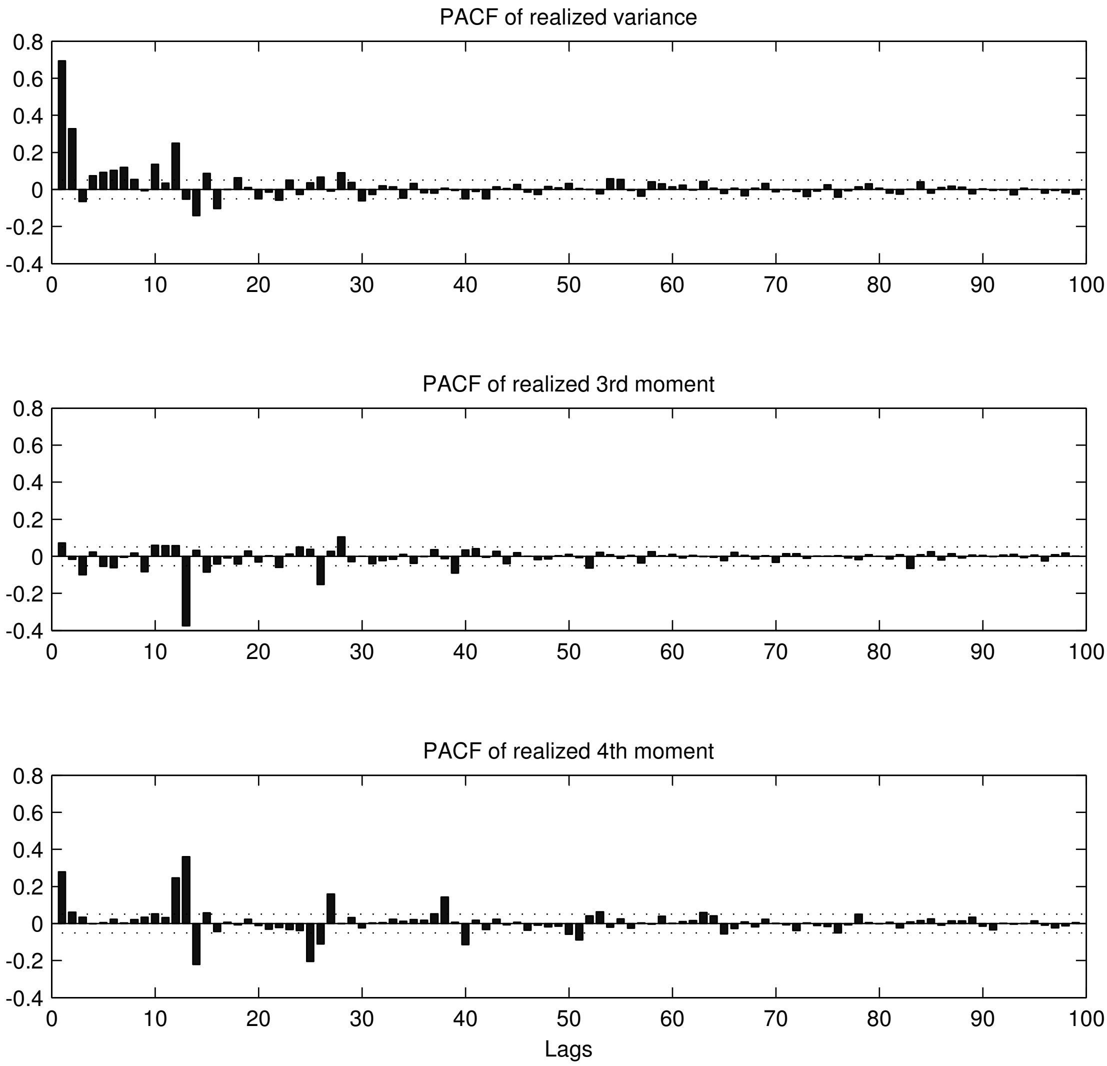

Figure B.2: Partial autocorrelation functions

The figure shows the partial autocorrelation function of realized variance, realized 3rd and realized 4th moment. Dotted lines define the 95% confidence interval.

Bibliography

AAS, K., HOBÆK HAFF, I., AND DIMAKOS, X. K. 2006. Risk estimation using the multivariate normal inverse Gaussian distribution. *Journal of Risk* 8:39–60.

ABAD, P. AND BENITO, S. 2013. A detailed comparison of value at risk estimates. *Mathematics and Computers in Simulation* 94:258–276.

ABAD, P., CHULIA, H., AND GOMEZ-PUIG, M. 2010. EMU and European government bond market integration. *Journal of Banking & Finance* 34:2851–2860.

ALEXANDER, C. 2010. Value-at-Risk Models, volume Vol. 4 of *Market risk analysis*. Wiley, Chichester.

AMAYA, D., CHRISTOFFERSEN, P., JACOBS, K., AND VASQUEZ, A. 2013. Does realized skewness predict the cross-section of equity returns? *working paper* .

AMMANN, M. AND REICH, C. 2001. VaR for nonlinear financial instruments - linear approximation or full Monte Carlo? *Financial Markets and Portfolio Management* 15:363–378.

ANDERSEN, T. G. AND BOLLERSLEV, T. 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39:885–905.

ANDERSEN, T. G., BOLLERSLEV, T., CHRISTOFFERSEN, P. F., AND DIEBOLD, F. X. 2006. Volatility and correlation forecasting, pp. 777–878. *In* G. Elliott, C. W. J. Granger, and A. Timmermann (eds.), Handbook of Economic Forecasting. North-Holland, Amsterdam.

ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X., AND EBENS, H. 2001. The distribution of realized stock return volatility. *Journal of Financial Economics* 61:43–76.

ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X., AND LABYS, P. 2003. Modeling and forecasting realized volatility. *Econometrica* 71:579–626.

ANDRESEN, A., KOEKEBAKKER, S., AND WESTGAARD, S. 2010. Modeling electricity forward prices using the multivariate normal inverse Gaussian distribution. *Journal of Energy Markets* 3:3–25.

APARICIO, F. M. AND ESTRADA, J. 2001. Empirical distributions of stock returns: European securities markets, 1990 - 95. *European Journal of Finance* 7:1–21.

ARTZNER, P. 1999. Coherent measures of risk. *Mathematical Finance* 9:203–228.

BAELE, L., FERRANDO, A., HÖRDAHL, P., KRYLOVA, E., AND MONNET, C. 2004. Measuring European financial integration. *Oxford Review of Economic Policy* 20:509–530.

BARNDORFF-NIELSEN, O. E. 1997. Normal inverse gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics* 24:1–13.

BARNDORFF-NIELSEN, O. E. AND SHEPHARD, N. G. 2002. Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B* 64:253–280.

BARONE-ADESI, G., GIANNOPOULOS, K., AND VOSPER, L. 1999. VaR without correlations for nonlinear portfolios. *Journal of Futures Markets* 19:583–602.

BARONE-ADESI, G., GIANNOPOULOS, K., AND VOSPER, L. 2002. Backtesting derivative portfolios with filtered historical simulation (FHS). *European Financial Management* 8:31–58.

BASEL COMMITTEE 2001. Operational risk. *Basel Committee on Banking Supervision* January 2001.

BASEL COMMITTEE 2006. International convergence of capital measurement and capital standards. *Basel Committee on Banking Supervision* June 2006.

BASEL COMMITTEE 2011. Basel III: A global regulatory framework for more resilient banks and banking systems. *Basel Committee on Banking Supervision* June 2011.

BASEL COMMITTEE 2012. Fundamental review of the trading book. *Basel Committee on Banking Supervision* May 2012.

BAUWENS, L., LAURENT, S., AND ROMBOUTS, J. V. K. 2006. Multivariate GARCH models: A survey. *Journal of Applied Econometrics* 21:79–109.

BEBER, A., BRANDT, M. W., AND KAVAJECZ, K. A. 2009. Flight-to-quality or flight-to-liquidity? Evidence from the Euro-area bond market. *Review of Financial Studies* 22:925–957.

BELTRATTI, A. AND MORANA, C. 2005. Statistical benefits of value-at-risk with long memory. *Journal of Risk* 7:21–45.

BENTH, F. E., CARTEA, A., AND KIESEL, R. 2008. Pricing forward contracts in power markets by the certainty equivalence principle: Explaining the sign of the market risk premium. *Journal of Banking & Finance* 32:2006–2021.

BENTH, F. E. AND KOEKEBAKKER, S. 2008. Stochastic modeling of financial electricity contracts. *Energy Economics* 30:1116–1157.

BENTH, F. E. AND MEYER-BRANDIS, T. 2009. The information premium for non-storable commodities. *Journal of Energy Markets* 2:111–140.

BERA, A. K. AND JARQUE, C. M. 1987. A test for normality of observations and regression residuals. *International Statistical Review* 55:163–172.

BERKOWITZ, J. AND O'BRIEN, J. 2002. How accurate are value-at-risk models at commercial banks? *Journal of Finance* 57:1093–1111.

BERNOTH, K., VON HAGEN, J., AND SCHUKNECHT, L. 2004. Sovereign risk premia in the European government bond market, volume 369 of *Working paper series / European Central Bank*. European Central Bank, Frankfurt am Main.

BESSEMBINDER, H., COUGHENOUR, J. F., SEGUIN, P. J., AND SMOLLER, M. M. 1996. Is there a term structure of futures volatilities? Reevaluating the Samuelson hypothesis. *Journal of Derivatives* 4:45–58.

BESSEMBINDER, H. AND LEMMON, M. L. 2002. Equilibrium pricing and optimal hedging in electricity forward markets. *Journal of Finance* 57:1347–1382.

BLACK, F. AND SCHOLES, M. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81:637–654.

BLATTBERG, R. C. AND GONEDES, N. J. 1974. A comparison of the stable and student distributions as statistical models for stock prices. *Journal of Business* 47:244–280.

BOLLERSLEV, T. P. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31:307–327.

BOLLERSLEV, T. P. 1987. A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* 69:542–547.

BOLLERSLEV, T. P. 2009. Glossary to ARCH (GARCH). *working paper* .

BOLLERSLEV, T. P., CHOU, R. Y., AND KRONER, K. F. 1992. ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics* 52:5–59.

BOND, S. A. AND PATEL, K. 2003. The conditional distribution of real estate returns: Are higher moments time varying? *Journal of Real Estate Finance and Economics* 26:319–339.

BOX, G. E. P. AND DRAPER, N. R. 1987. Empirical Model-Building and Response Surfaces. Wiley, New York.

BROOKS, C., BURKE, S. P., HERAVI, S., AND PERSAND, G. 2005. Autoregressive conditional kurtosis. *Journal of Financial Econometrics* 3:399–421.

BYSTRÖM, H. N. E. 2003. The hedging performance of electricity futures on the Nordic power exchange. *Applied Economics* 35:1–11.

BYSTRÖM, H. N. E. 2005. Extreme value theory and extremely large electricity price changes. *International Review of Economics & Finance* 14:41–55.

CAPPIELLO, L., ENGLE, R. F., AND SHEPPARD, K. 2003. Asymmetric dynamics in the correlations of global equity and bond returns, volume 204 of *Working paper series / European Central Bank*. European Central Bank, Frankfurt am Main.

CHAN, K. F. AND GRAY, P. 2006. Using extreme value theory to measure value-at-risk for daily electricity spot prices. *International Journal of Forecasting* 22:283–300.

CHEN, Y., HÄRDLE, W., AND JEONG, S.-O. 2005. Nonparametric risk management with generalized hyperbolic distributions. *Journal of the American Statistical Association* 103:910–923.

CHEN, Y. AND LU, J. 2012. Value at risk estimation. *In* J.-C. Duan, W. Härdle, and J. E. Gentle (eds.), Handbook of Computational Finance, Springer Handbooks of Computational Statistics. Springer, Heidelberg and New York.

CHERUBINI, U., LUCIANO, E., AND VECCHIATO, W. 2004. Copula Methods in Finance. Wiley, Chichester.

CHOW, G. C. 1960. Test of equality between sets of coefficients in two linear regressions. *Econometrica* 28:591–605.

CHRISTIANSEN, C. 2007. Volatility-spillover effects in European bond markets. *European Financial Management* 13:921–946.

CHRISTOFFERSEN, P. F. 2003. Elements of Financial Risk Management. Academic Press, Amsterdam and Boston.

CLEMENTS, M. P., GALVÃO, A. B., AND KIM, J. H. 2008. Quantile forecasts of daily exchange rate returns from forecasts of realized volatility. *Journal of Empirical Finance* 15:729–750.

CODOGNO, L., FAVERO, C., MISSALE, A., PORTES, R., AND THUM, M. 2003. Yield spreads on EMU government bonds. *Economic Policy* 18:503–532.

CONRAD, J. S. AND KAUL, G. 1988. Time-variation in expected returns. *Journal of Business* 61:409–425.

CONT, R. 2001. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance* 1:223–236.

CORNISH, E. A. AND FISHER, R. A. 1937. Moments and cumulants in the specification of distributions. *Extrait de la Revue de l'Institute International* 4:1–14.

CORSI, F. 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7:174–196.

COTTIN, C. AND DÖHLER, S. 2009. Risikoanalyse: Modellierung, Beurteilung und Management von Risiken mit Praxisbeispielen. Vieweg & Teubner, Wiesbaden.

CREMERS, H., MEHMKE, F., AND PACKHAM, N. 2012. Validierung von Konzepten zur Messung des Marktrisikos: Insbesondere des Value at Risk und des Expected Shortfall. *working paper series, Frankfurt School of Finance & Management, No. 192* .

DOWD, K. 2005. Measuring Market Risk. Wiley, Chichester, 2nd edition.

DUMOUCHEL, W. H. 1983. Estimating the stable index alpha in order to measure tail thickness: A critique. *The Annals of Statistics* 11:1019–1031.

DUONG, H. N. AND KALEV, P. S. 2008. The Samuelson hypothesis in futures markets: An analysis using intraday data. *Journal of Banking & Finance* 32:489–500.

EDERINGTON, L. H. AND GUAN, W. 2010. Longer-term time-series volatility forecasts. *Journal of Financial and Quantitative Analysis* 45:1055–1076.

ENGLE, R. F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50:987–1007.

ENGLE, R. F. 2002a. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics* 20:339–350.

ENGLE, R. F. 2002b. New frontiers for ARCH models. *Journal of Applied Econometrics* 17:425–446.

ENGLE, R. F. AND BOLLERSLEV, T. P. 1986. Modelling the persistence of conditional variances. *Econometric Reviews* 5:1–50.

ENGLE, R. F. AND MANGANELLI, S. 2004. CAViaR: Conditional autoregressive value-at-risk by regression quantiles. *Journal of Business & Economic Statistics* 22:367–381.

ENGLE, R. F. AND NG, V. K. 1993. Measuring and testing the impact of news on volatility. *Journal of Finance* 48:1949–1987.

ERIKSSON, A., GHYSELS, E., AND WANG, F. 2009. The normal inverse Gaussian distribution and the pricing of derivatives. *Journal of Derivatives* 16:23–37.

FAMA, E. F. 1965. The behavior of stock-market prices. *Journal of Business* 38:34–105.

FITE, D. AND PFLEIDERER, P. 1995. Should firms use derivatives to manage risks?, pp. 139–169. *In* W. Beaver and G. Parker (eds.), Risk Management: Problems and Solutions. McGraw-Hill, New York.

FLETEN, S.-E. AND LEMMING, J. 2003. Constructing forward price curves in electricity markets. *Energy Economics* 25:409–424.

FORSBERG, L. AND BOLLERSLEV, T. 2002. Bridging the gap between the distribution of realized (ECU) volatility and ARCH modelling (of the Euro): The GARCH-NIG model. *Journal of Applied Econometrics* 17:535–548.

FRESTAD, D., BENTH, F. E., AND KOEKEBAKKER, S. 2010. Modeling term structure dynamics in the nordic electricity swap market. *Energy Journal* 31:53–86.

GABRIEL, C. AND LAU, C. 2014. On the distribution of government bond returns: Evidence from the EMU (Original copyright notice – With permission of Springer Science+Business Media). *Financial Markets and Portfolio Management* 28:181–203.

GARCIA, R. C. 2005. A GARCH forecasting model to predict day-ahead electricity prices. *Institute of Electrical and Electronics Engineers* 20:867–874.

GENÇAY, R. AND SELÇUK, F. 2004. Extreme value theory and value-at-risk: Relative performance in emerging markets. *International Journal of Forecasting* 20:287–303.

GIANNOPOULOS, K. AND TUNARU, R. 2005. Coherent risk measures under filtered historical simulation. *Journal of Banking & Finance* 29:979–996.

GILLI, M. 2006. An application of extreme value theory for measuring financial risk. *Computational Economics* 27:207–228.

GIOT, P. AND LAURENT, S. 2004. Modelling daily value-at-risk using realized volatility and ARCH type models. *Journal of Empirical Finance* 11:379–398.

GLASSERMAN, P. 2004. Monte Carlo Methods in Financial Engineering. Springer, New York.

GLOSTEN, L. R., JAGANNATHAN, R., AND RUNKLE, D. E. 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48:1779–1801.

GOMEZ-PUIG, M. 2009. Systemic and idiosyncratic risk in EU-15 sovereign yield spreads after seven years of monetary union. *European Financial Management* 15:971–1000.

GREENE, W. H. 2003. Econometric Analysis. Pearson Prentice Hall, Upper Saddle River, NJ, 5th edition.

HANSEN, B. E. 1994. Autoregressive conditional density estimation. *International Economic Review* 35:705–730.

HANSEN, P. R., HUANG, Z., AND SHEK, H. H. 2012. Realized GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics* 27:877–906.

HARTMANN, P., MADDALONI, A., AND MANGANELLI, S. 2003. The Euro-area financial system: Structure, integration, and policy initiatives. *Oxford Review of Economic Policy* 19:180–213.

HENDRICKS, D. 1996. Evaluation of value-at-risk models using historical data. *Federal Reserve Bank of New York Economic Policy Review* 2:39–69.

HOLTON, G. A. 2002. History of value-at-risk: 1922-1998. *working paper* .

HOLTON, G. A. 2003. Value-at-Risk: Theory and Practice. Academic Press, Amsterdam.

HULL, J. AND WHITE, A. 1998. Incorporating volatility updating into the historical simulation method for value-at-risk. *Journal of Risk* 1:5–19.

JEGADEESH, N. AND TITMAN, S. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* 48:65–91.

JORION, P. 2007. Value at Risk: The New Benchmark for Managing Financial Risk. McGraw-Hill, New York, 3rd edition.

J.P. MORGAN AND REUTERS 1996. RiskMetrics – Technical Document. J. P. Morgan, New York NY, 4th edition.

KALEMANOVA, A., SCHMID, B., AND WERNER, R. 2007. The normal inverse Gaussian distribution for synthetic CDO pricing. *Journal of Derivatives* 14:80–93.

KARLIS, D. 2002. An EM type algorithm for maximum likelihood estimation of the normal–inverse Gaussian distribution. *Statistics & Probability Letters* 57:43–52.

KIM, Y. S., RACEV, S. T., BIANCHI, M. L., AND FABOZZI, F. J. 2010. Tempered stable and tempered infinitely divisible GARCH models. *Journal of Banking & Finance* 34:2096–2109.

KON, S. J. 1984. Models of stock returns: A comparison. *Journal of Finance* 39:147–165.

KUESTER, K., MITTNIK, S., AND PAOLELLA, M. S. 2006. Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics* 4:53–89.

KWON, K.-Y. AND KISH, R. J. 2002. Technical trading strategies and return predictability: NYSE. *Applied Financial Economics* 12:639–653.

LAITENBERGER, J. AND LAU, C. 2014. Measuring risk in electricity forward returns. *working paper* .

LAOPODIS, N.-T. 2008. Government bond market integration within European Union. *International Research Journal of Finance and Economics* 19:56–76.

LAU, C. 2015. A simple normal inverse gaussian-type approach to calculate value at risk based on realized moments. *Journal of Risk* 17:1–18.

LILLESTØL, J. 2000. Risk analysis and the NIG distribution. *Journal of Risk* 2:41–56.

LILLIEFORS, H. W. 1967. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association* 62:399–402.

LIMA, L. R. AND DE NÉRI, B. A. P. 2007. Comparing value-at-risk methodologies. *Brazilian Review of Econometrics* 27:1–25.

LING, S. AND MCALEER, M. 2003. Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory* 19:280–310.

LIU, P. 2012. Volatility, duration and value-at-risk. PhD thesis, University of Western Ontario – Electronic Thesis and Dissertation Repository. Paper 933.

LO, A. W. AND MACKINLAY, A. C. 1989. The size and power of the variance ratio test in finite samples: A Monte Carlo investigation. *Journal of Econometrics* 40:203–238.

LONGIN, F. M. 1996. The asymptotic distribution of extreme stock market returns. *Journal of Business* 69:383–408.

LONGSTAFF, F. A. AND WANG, A. W. 2004. Electricity forward prices: A high-frequency empirical analysis. *Journal of Finance* 59:1877–1900.

MANDELBROT, B. 1963. The variation of certain speculative prices. *Journal of Business* 36:394–419.

MANGANELLI, S. 2002. DQtest.

MARINELLI, C., D'ADDONA, S., AND RACEV, S. T. 2007. A comparison of some univariate models for value-at-risk and expected shortfall. *International journal of theoretical and applied finance* 10:1043–1075.

MARKOWITZ, H. 1952. Portfolio selection. *Journal of Finance* 7:77–91.

MCALEER, M., CHAN, F., AND MARINOVA, D. 2007. An econometric analysis of asymmetric volatility: Theory and application to patents. *Journal of Econometrics* 139:259–284.

MCNEIL, A. J. 1999. Extreme value theory for risk managers, pp. 93–113. *In* Risk Books (ed.), Internal Modelling and CAD II. Risk Books, London.

MCNEIL, A. J. AND FREY, R. 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance* 7:271–300.

MCNEIL, A. J., FREY, R., AND EMBRECHTS, P. 2005. Quantitative Risk Management: Concepts, Techniques and Tools. Princeton University Press, New Jersey.

NELSON, D. B. 1991. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59:347–370.

NG, V. K. AND PIRRONG, S. C. 1994. Fundamentals and volatility: Storage, spreads, and the dynamics of metals prices. *Journal of Business* 67:203–230.

NOLAN, J. P. 2001. Maximum likelihood estimation and diagnostics for stable distributions. *In* O. E. Barndorff-Nielsen, T. Mikosch, and S. I. Resnick (eds.), Lévy Processes: Theory and Applications. Birkhäuser, Boston.

PAGANO, M. AND VON THADDEN, E.-L. 2004. The European bond markets under EMU. *Oxford Review of Economic Policy* 20:531–554.

PAOLELLA, M. S. 2007. Intermediate Probability: A Computational Approach. Wiley, Chichester.

PEIRÓ, A. 1994. The distribution of stock returns: International evidence. *Applied Financial Economics* 4:431–439.

PEIRÓ, A. 1999. Skewness in financial returns. *Journal of Banking & Finance* 23:847–862.

PÉRIGNON, C. AND SMITH, D. R. 2008. A new approach to comparing VaR estimation methods. *Journal of Derivatives* 16:54–66.

PIRRONG, C. AND JERMAKYAN, M. 1999. Valuing power and weather derivatives on a mesh using finite difference methods, pp. 59–69. *In* Risk Books (ed.), Energy Modeling and the Management of Uncertainty. Risk Books, London.

PRAETZ, P. D. 1972. The distribution of share price changes. *Journal of Business* 45:49–55.

PRESS, J. 1967. A compound events model for security prices. *Journal of Business* 40:317–335.

PRITSKER, M. 2006. The hidden dangers of historical simulation. *Journal of Banking & Finance* 30:561–582.

RACHEV, S. T. AND MITTNIK, S. 2000. Stable Paretian Models in Finance. Series in financial economics and quantitative analysis. Wiley, New York.

RACHEV, S. T., RACHEVA-IOTOVA, B., AND STOYANOV, S. 2010. Capturing fat tails. *Risk* May:76–80.

RACHEV, S. T., SCHWARTZ, E. S., AND KHINDANOVA, I. 2003. Stable modeling of market and credit value at risk, pp. 249–328. *In* S. T. Racev (ed.), Handbook of Heavy Tailed Distributions in Finance. Elsevier, Amsterdam.

SAMUELSON, P. A. 1965. Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review* 6:41–49.

SCARROTT, C. AND MACDONALD, A. 2012. A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT – Statistical Journal* 10:33–60.

SGHERRI, S. AND ZOLI, E. 2009. Euro area sovereign risk during the crisis: "European Department.", volume WP/09/222 of *IMF working paper*. International Monetary Fund European Dept., Washington, D.C.

SHAWKY, H. A., MARATHE, A., AND BARRETT, C. L. 2003. A first look at the empirical relation between spot and futures electricity prices in the United States. *Journal of Futures Markets* 23:931–955.

SHEPHARD, N. AND SHEPPARD, K. 2010. Realising the future: Forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics* 25:197–231.

SHEPHARD, N. G. 1996. Statistical aspects of ARCH and stochastic volatility, pp. 1–68. *In* D. R. Cox, D. V. Hinkley, and O. E. Barndorff-Nielsen (eds.), Time Series Models in Econometrics, Finance and other Fields, volume 65 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, London.

SHEPPARD, K. 2013. MFE Toolbox.

SKLAR, A. 1959. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris* 8:229–231.

SOLIBAKKE, P. B. 2010. Corporate risk management in European energy markets. *Journal of Energy Markets* 3:93–131.

TALEB, N. N. 2007. The Black Swan: The Impact of the Highly Improbable. Random House, New York.

TURTLE, H. J., BUSE, A., AND KORKIE, R. M. 1994. Tests of conditional asset pricing with time-varying moments and risk prices. *Journal of Financial and Quantitative Analysis* 29:15–29.

VEHVILÄINEN, I. 2002. Basics of electricity derivative pricing in competitive markets. *Applied Mathematical Finance* 9:45–60.

VENTER, J. H. AND DE JONGH, P. J. 2002. Risk estimation using the normal inverse Gaussian distribution. *Journal of Risk* 4:1–23.

WAGNER, F. 2000. Risk Management im Erstversicherungsunternehmen: Modelle - Strategien - Ziele - Mittel. Verlag Versicherungswirtschaft, Karlsruhe.

WERNER, R. 2004. Normal Inverse Gaussion Distribution.

WERON, R. 2009. Heavy-tails and regime-switching in electricity prices. *Mathematical Methods of Operations Research* 69:457–473.

YOUNG, M. AND GRAFF, R. 1995. Real estate is not normal: A fresh look at real estate return distributions. *Journal of Real Estate Finance and Economics* 10:225–259.

ZHANG, L., MYKLAND, P. A., AND AÏT-SAHALIA, Y. 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100:1394–1411.