Modeling Asymmetries in Financial Data with Multiplicative Error Models

Jarkko Miettinen

Academic dissertation to be presented,
by the permission of the Faculty of Social Sciences of the University of Helsinki,
for public examination in the lecture room PIII, Porthania, Yliopistonkatu 3,
on December 19, 2009 at 10.00.

Helsinki 2009

Jarkko Miettinen

Modeling Asymmetries in Financial Data with Multiplicative Error Models

Acknowledgements

This thesis was written while working as an Assistant at the Department of Mathematics and Statistics, University of Helsinki, between the years 2004 and 2009. The author would like to thank the Department for providing first class facilities for this research.

I am deeply indebted to my supervisor, Professor Pentti Saikkonen, for his most generous use of his time and effort during the process of writing this thesis. Most of all, I would like to thank him for suggesting I do postgraduate studies and pointing out research problems for Chapters 2 and 3 of this work. As usual, the author is responsible of all the shortcomings of the thesis.

I am grateful to the pre-examiners of this thesis, Dr. Markus Haas and Professor Seppo Pynnönen, for their comments and detailed work. I have also gained substantially from the comments of the examiners of my Licentiate thesis, Professor Markku Lanne and University Teacher, Dr. Pekka Pere. Professor Risto Lehtonen was most helpful with administrative tasks in the final stages of the research.

My warmest thanks go to Assistant Professor Alexandru Badescu for introducing and letting me to participate in his research project. Working with him provided me with new points of view on financial econometrics. I would also like to thank all those numerous people who have helped me during my postgraduate studies both at the Department and elsewhere, but whose names I have forgotten to mention.

Finally, my humble thanks are due to my parents and to my sister for their encouragement, patience and support during the past years.

Helsinki, October 2009

Jarkko Miettinen

Abstract

This thesis addresses modeling of financial time series, especially stock market returns and daily price ranges. Modeling data of this kind can be approached with so-called multiplicative error models (MEM). These models nest several well known time series models such as GARCH, ACD and CARR models. They are able to capture many well established features of financial time series including volatility clustering and leptokurtosis.

In contrast to these phenomena, different kinds of asymmetries have received relatively little attention in the existing literature. In this thesis asymmetries arise from various sources. They are observed in both conditional and unconditional distributions, for variables with non-negative values and for variables that have values on the real line. In the multivariate context asymmetries can be observed in the marginal distributions as well as in the relationships of the variables modeled. New methods for all these cases are proposed.

Chapter 2 considers GARCH models and modeling of returns of two stock market indices. The chapter introduces the so-called generalized hyperbolic (GH) GARCH model to account for asymmetries in both conditional and unconditional distribution. In particular, two special cases of the GARCH-GH model which describe the data most accurately are proposed. They are found to improve the fit of the model when compared to symmetric GARCH models. The advantages of accounting for asymmetries are also observed through Value-at-Risk applications.

Both theoretical and empirical contributions are provided in Chapter 3 of the thesis. In this chapter the so-called mixture conditional autoregressive range (MCARR) model is introduced, examined and applied to daily price ranges of the Hang Seng Index. The conditions for the strict and weak stationarity of the model as well as an expression for the autocorrelation function are obtained by writing the MCARR model as a first order autoregressive process with random coefficients. The chapter also introduces inverse gamma (IG) distribution to CARR models. The advantages of CARR-IG and MCARR-IG specifications over conventional CARR models are found in the empirical application both in- and out-of-sample.

Chapter 4 discusses the simultaneous modeling of absolute returns and daily price ranges. In this part of the thesis a vector multiplicative error model (VMEM) with asymmetric Gumbel copula is found to provide substantial benefits over the existing VMEM models based on elliptical copulas. The proposed specification is able to capture the highly asymmetric dependence of the modeled variables thereby improving the performance of the model considerably. The economic significance of the results obtained is established when the information content of the volatility forecasts derived is examined.

Tiivistelmä

Tässä väitöskirjassa tarkastellaan rahoitusaikasarjojen, erityisesti osakkeiden tuottojen ja osakkeiden hintojen päivittäisen vaihteluvälin, mallintamista. Tämän tyyppisten aineistojen mallintamisessa voidaan käyttää niin sanottuja multiplikatiivisen virhetermin malleja (MEM). Näihin malleihin kuuluvat monet tunnetut aikasarjamallit, kuten GARCH-, ACD- ja CARR-mallit. Niillä voidaan mallintaa useita rahoitusaikasarjalle tyypillisiä piirteitä, kuten volatiliteetin klusteroitumista sekä jakauman paksuhäntäisyyttä.

Vastoin kuin edellä mainittuja ilmiöitä, on erilaisia aineistossa havaittavia epäsymmetrisyyksiä käsitelty aikaisemmassa kirjallisuudessa verraten vähän. Tutkimuksessa tarkasteltavia epäsymmetrisyyksiä havaitaan aineistossa sekä ehdollisessa että ehdottomassa jakaumassa ja niin ei-negatiivisilla muuttujilla kuin muuttujilla, jotka saavat arvoja koko reaaliakselillakin. Moniulotteisessa tapauksessa epäsymmetrisyyksiä on reunajakaumissa ja muuttujien välisissä riippuvuuksissa. Väitöskirjassa esitetään uusia menetelmiä kaikkiin edellä mainittuihin tapauksiin.

Luku 2 käsittelee GARCH-malleja ja osakeindeksien tuottojen mallintamista. Aineiston epäsymmetrisyyksien huomioimiseksi luvussa esitellään niin sanottu yleistetty hyperbolinen GARCH-malli. Mallista voidaan erottaa kaksi erikoistapausta, jotka eniten parantavat mallin sopivuutta symmetrisiin vaihtoehtoihin verrattuna. Epäsymmetrisyyden huomiointi osoittautuu edulliseksi myös niin sanotuissa Value-at-Risk sovelluksissa.

Luvun 3 aiheena on ehdollisen autoregressiivisen vaihteluvälin sekoitusmalli (MCARR). Mallille esitetään ehdot vahvalle ja heikolle stationaarisuudelle ja johdetaan autokorrelaatiofunktio kirjoittamalla malli uudelleen moniulotteisena ensimmäisen kertaluvun satunnaiskertoimisena autoregressiivisenä prosessina. Luvussa ehdotetaan myös käänteisen gammajakauman (IG) käyttöä CARR-mallinnuksessa. CARR-IG- ja MCARR-IG-mallien edut aikaisemmin ehdotettuihin vaihtoehtoihin verrattuna tulevat esille sovelluksessa Hang Seng -indeksin vaihteluväliin sekä estimointi- että ennusteperiodilla.

Neljännessä luvussa mallinnetaan samanaikaisesti tuottojen itseisarvoa ja hin-

tojen päivittäistä vaihteluväliä. Luvussa todetaan epäsymmetriseen Gumbel-kopulaan perustuvan vektorimultiplikatiivisen virhetermin mallin (VMEM) tarjoavan huomattavia etuja verrattaessa sitä aikaisemmin esiteltyihin elliptisiin kopuloihin perustuviin VMEM-malleihin. Ehdotetulla spesifikaatiolla voidaan ottaa huomioon mallinnettavien muuttujien vahvasti epäsymmetrinen riippuvuus. Näin ollen mallin sopivuus aineistoon paranee huomattavasti. Viitteitä saatujen tulosten taloudellisesta merkitsevyydestä nähdään tarkasteltaessa eri malleista johdettujen volatiliteettiennusteiden informaatiosisältöä.

Contents

3 Mixture MEM models and conditional

autoregressive range   56

4 Modeling financial volatility measures with a VMEM model based

on an asymmetric copula   85

1 Introduction

As defined by Engle (2001, 53), financial econometrics is simply the application of econometric tools to financial data. While this area of research is today one of the most active in econometrics (see Bollerslev 2001, 41-42), such has not always been the case. Indeed, until the beginning of the eighties most papers in empirical finance relied on fairly simplistic data analytical tools. However, since then the developments have been rapid. Advances have included acceleration of computing power, increased availability of data for financial instruments, and the development of more sophisticated econometric techniques.

Many financial time series, such as stock returns and exchange rate returns, are characterized by two so-called stylized facts: excess kurtosis of the unconditional distribution and volatility clustering. These facts can be described with ARCH type time series models proposed by Engle (1982) and Bollerslev (1986). For a survey of these models see, among others Bollerslev et al. (1992), Bera and Higgins (1993), Bollerslev et al. (1994), Palm (1996), Degiannakis and Xekalaki (2004), Diebold (2004), and Bollerslev (2008). As argued by Engle (2002b, 425-426), univariate ARCH/GARCH models are nowadays a well researched area as the literature on different specifications, theorems for autocorrelations, moments and stationarity and ergodicity, and Value-at-Risk applications is extensive. However, several new areas of research have been established and financial econometrics can be expected to develop rapidly in the future. Engle (2002b, 444) mentions five areas for future work: high frequency volatility, multivariate models, derivatives pricing, modeling non-negative processes, and analyzing conditional simulations by Least Squares Monte Carlo. Because the methods proposed in this thesis are related to the first four of these topics, we will discuss them in more detail in subchapter 1.2. First, an outline of the thesis is presented.

1.1 Outline of the thesis

The thesis consists of three actual chapters each concerning the modeling of financial data from various perspectives. In the second chapter of the thesis we develop

GARCH models with generalized hyperbolic (GH) innovations. This chapter is divided into four subchapters, the first of which introduces GH distributions. In the second subchapter GARCH models based on these distributions are proposed. The third subchapter presents an application to US and European data while the fourth subchapter concludes with a discussion.

In the first part of Chapter 3, we introduce an autoregressive model with random coefficients. In subchapters 3.2 and 3.3 so-called mixture multiplicative error models (MMEM) are defined and represented as a special case of the model discussed in subchapter 3.1. Based on this representation, conditions for both strict and weak stationarity of the MMEM models are given. In subchapter 3.4 the autocorrelation function of the considered model is derived, and an application to Asian data is presented in subchapter 3.5. Chapter 3 ends with a discussion.

In the last chapter of the thesis vector multiplicative error models (VMEM) are examined. First we introduce copulas in subchapter 4.1 and construct VMEM models based on them in subchapter 4.2. An application to US data is found in subchapter 4.3 and a discussion in subchapter 4.4.

The rest of the introduction is organized as follows. As already mentioned, in subchapter 1.2 relevant topics in financial econometrics, such as multivariate models and modeling of intraday data, are discussed. In subchapters 1.3, 1.4 and 1.5 more detailed introductions to Chapters 2, 3 and 4 are presented. Subchapter 1.6 includes conclusions. References are given at the end of each main chapter. Suggestions for future work are given in the discussions at the end of Chapters 2, 3, and 4.

1.2 Some current topics in financial econometrics

1.2.1 High frequency volatility models

A natural extension of the volatility models introduced and applied to daily, weekly or monthly data are models for data within the day. Early models (see Engle 2002b, 427 and references therein) were introduced for regularly spaced data and focused on the so-called time of the day effect. However, as discussed e.g. by Engle (2000, 1-2), ultimately it is desirable to model irregularly spaced data. Data for which all

transactions are recorded is called 'ultra-high frequency' data.

When data arrive at random dates and these times themselves carry information, the basic procedure is to model the associated variables called marks conditional on times, and then model the times separately (Engle 2000, 21). In his application Engle (2000) uses the Autoregressive Conditional Duration (ACD) model of Engle and Russel (1998) to model arrival times of trades. A semiparametric approach is applied to the estimation of the hazard function. Finally, price quotes are examined to obtain models of volatility conditional on transaction times. Evidence is found that longer durations and longer expected durations are associated with lower volatilities. Moreover, higher bid-ask spreads and larger volumes both predict rising volatility.

The approach proposed by Engle (2000) was extended by Manganelli (2005) by elaborating a system where returns and volatilities directly interact with duration and volume. This is accomplished by first modeling volumes with a model similar to the ACD model. Then duration, volume and returns are modeled simultaneously with a special type of vector autoregression. The system developed allows causal and feedback effects among these variables. The main findings include trading clustering and different behavior of frequently and infrequently traded stocks. As a direction for future research Manganelli (2005, 398) suggests adding other variables, such as depth and spread, to the model. Also, a study exploring the relationships between different markets is suggested.

In addition, better daily models have been constructed using intra-daily data. Models based upon so-called 'realized volatility'[1] are built by Andersen et al. (2001a, 2001b, 2003) and further developed e.g. by Lanne (2006).

1.2.2   Multivariate models

An obvious generalization of univariate ARCH/GARCH models are multivariate volatility models. Surprisingly, even though the literature on volatility models is extensive, only a small fraction of it is devoted to multivariate GARCH models.

---

[1]We simply define that the daily realized volatility RV_t at time t is given by $RV_t = \sum_{i=1}^{\tau}(\log P_{t,i} - \log P_{t,i-1})^2$, where $P_{t,i}$ is the price for the time $i \times \Delta$ in the trading day t, and Δ is the time interval (there are τ equal-length intervals divided in a trading day) (Chou et al. 2008).

Most notable studies on these models include multivariate GARCH-M by Bollerslev et al. (1988), constant conditional correlation GARCH by Bollerslev (1990), vec and BEKK-GARCH by Engle and Kroner (1995) and dynamic conditional correlation GARCH by Engle (2002a).

As noted by Engle (2002a, 339), in very few articles have more than five assets been considered. The most successful model for these cases has been the CCC model by Bollerslev (1990) because of its computational simplicity (Engle 2004, 418). In most cases the number of parameters is too large for easy estimation. As a solution, several factor and orthogonal models have been introduced in the literature. Examples include Engle et al. (1990) and Lanne and Saikkonen (2007a). So-called copula-GARCH models, in which the conditional dependence is modeled using copula functions (see subchapter 4.1 for a discussion), were recently proposed by Jondeau and Rockinger (2006) and Patton (2006). For a survey and proposals for future work on multivariate GARCH models we refer to Bauwens et al. (2006) and Silvennoinen and Teräsvirta (2008).

1.2.3 Options pricing

One of the most interesting areas of research in financial econometrics at the moment is pricing of options when the underlying asset follows a GARCH process. Options pricing in general has been an important area of research since the work of Black and Scholes (1973) and Merton (1973), even though the history of option pricing theory can be traced back to Bachelier (1900). However, because the standard Black-Scholes model does not take the well-documented heteroskedasticity of the asset returns into account, several alternative option pricing models have been proposed (see Duan 1995, 13 for references). In general, the proposed continuous time options pricing models face the difficulty that the variance rate is not observable. By contrast, discrete-time models, such as GARCH models, have the advantage of the relative ease of their estimation and possibility of diagnostic model checking.

It can be argued that research problems with GARCH option pricing arise from two sources. The first problem is the correct specification of the return dynamics and

the form of the return distribution. The second is finding an appropriate approach to risk neutralization. This problem derives from the fact that under GARCH models markets are incomplete[2] and, consequently, there exists an infinite number of risk neutral measures that can be used to construct option prices.

In the seminal paper of GARCH option pricing, Duan (1995) uses the local risk neutral valuation principle (LRNVR). The choice is justified by the argument that the representative agent in an economy is an expected utility maximizer, and the utility is additive and time-separable (Badescu and Kulperger 2008, 70). Other assumptions made by Duan (1995) include conditional normal distribution of asset returns and the invariance of the conditional volatility to the change of measure. However, in Duan et al. (2006) the so-called stochastic discount factor is considered for constructing a risk-neutralized dynamic for jump GARCH model, whereas the Esscher transform is employed by Siu et al. (2004), Elliot et al. (2006), Christoffersen et al. (2006) and Badescu and Kulperger (2008). The impact of the choice of the risk-neutralization method is examined by Badescu et al. (2008), who use a generalized local risk-neutral valuation relationship, an Esscher transform, and an extended Girsanov principle coupled with the mixture GARCH model first suggested by Haas et al. (2004) and Alexander and Lazar (2006). In their empirical application to Standard & Poor's 500 European Call option pricing Badescu et al. (2008) show that allowing the volatility model to capture the skewness and leptokurtosis of the data improves the performance of the option pricing model for both short and long maturity options compared to the case where the underlying stock dynamics are modeled with an asymmetric GARCH model. In their proposal for future work, option pricing using the type of methods discussed and developed in the second chapter of the thesis is mentioned. A study on option pricing with generalized hyperbolic innovations (see discussion in subchapter 1.3) is also considered by Chorro et al. (2008a, 2008b).

---

[2]In complete markets every contingent claim is perfectly replicable by a self financing porfolio.

1.2.4 Multiplicative error models for non-negative processes

In financial time series analysis the problem is often to model non-negative valued processes. This occurs when considering variables such as volumes, trades, durations, realized volatility, daily price range, and so on. Non-negativity has traditionally been approached by two methods: the non-negativity has been ignored or logs have been taken. The disadvantages of these approaches have been discussed by Engle (2002b, 428-429). As a solution, the so-called multiplicative error model (MEM) is proposed. The MEM model is formulated such that the non-negativity of the process is automatically satisfied. We shall discuss these models in more detail in subchapter 1.4 and in Chapter 3. It can briefly be mentioned that the MEM model nests the (squared) GARCH model of Bollerslev (1986), the ACD model of Engle and Russel (1998) as well as the conditional autoregressive range (CARR) model of Chou (2005). Engle (2002b, 432-433) illustrated how the MEM model is applied to realized volatility. A vector MEM model (VMEM) is also introduced by Engle (2002b). This model is applied to absolute returns, daily price range and realized volatility by Engle and Gallo (2006). More recently, the VMEM model has been investigated by Cipollini et al. (2006, 2007, 2008). We shall discuss VMEM models more specifically in subchapter 1.5 and in Chapter 4.

Potential applications of (V)MEM models are wide-ranging. Topics for further research include, for example, specification of the conditional distribution, time varying densities, more general model specifications, and forecasting performance.[3]

Of these topics, this thesis contributes to modeling non-negative processes as well as multivariate models. As already stated, we suggest a mixture multiplicative error model (MMEM) for the daily price range. Our contribution to multivariate models concerns VMEM models. We discuss these models in more detail in subchapter 1.5. The present thesis also addresses asymmetric univariate GARCH models which are most likely to prove relevant in the options pricing in the future. This research problem will be introduced in the next subchapter.

---

[3]For a recent empirical work see e.g. Engle et al. (2008).

1.3 GARCH modeling with generalized hyperbolic distributions

The GARCH-in-Mean (GARCH-M) model, originally introduced by Engle et al. (1987), allows the mean of the series to depend on its volatility. At this point the model is specified in a general form as

$$r_t = m_t + h_t^{1/2}\eta_t, \tag{1}$$

where η_t is a sequence of independent, identically distributed (*i.i.d.*) random variables with zero mean and unit variance, $h_t^{1/2}$ is a (positive) volatility process which describes the conditional heteroskedasticity in the observed process r_t, and m_t is a (stationary) process which describes the conditional mean of r_t. In the model h_{t-j} $(j > 0)$ and η_t are assumed to be independent. In the GARCH-M model the conditional mean is assumed to be a function of the conditional variance h_t. For instance, a positive relation between the excess return on the stock market and the conditional variance can be motivated by Merton's (1973) Intertemporal Capital Asset Pricing Model.

In this thesis, especially in the second chapter, the primary interest is in the distribution of the error term η_t. Engle (1982) used the normal distribution but, in order to fully capture the observed excess kurtosis, more fat-tailed distributions were proposed in subsequent literature. The best known alternative distributions are the t distribution (Bollerslev 1987) and the generalized error distribution (Nelson 1991). Both of these distributions are symmetric.

In the case of equity returns, both unconditional and conditional skewness have been empirically observed. Theoretically conditional skewness can be explained, for example, by the so-called volatility feedback effect and leverage effect (see Lanne and Saikkonen 2007b, Campbell and Hentschel 1992 and Christie 1982). This asymmetry can be approached either by specifying the conditional variance equation in an asymmetric way or by allowing for an asymmetric conditional distribution. Probably the best-known proposals of the former approach are the EGARCH model by

Nelson (1991) and the GJR-GARCH model by Glosten et al. (1993). One of the first papers introducing a skewed conditional distribution to GARCH modeling is Hansen (1994), where a skewed Student's t distribution with time varying shape parameters is considered. The so-called normal inverse Gaussian (NIG) distribution was introduced to financial econometrics by Barndorff-Nielsen (1997)[4] and first applied to GARCH models by Andersson (2001), Jensen and Lunde (2001) and Forsberg and Bollerslev (2002). Haas et al. (2004) and Alexander and Lazar (2006) consider a combination of mixed normal distributions and a GARCH-type dynamic structure. A distribution we call the normal reciprocal gamma (NRG) distribution was introduced to the modeling of financial data by Aas and Haff (2006), who called it the generalized hyperbolic (GH) skew Student's t-distribution. The Fractionally Integrated GARCH (FIGARCH) model was generalized by Kiliç (2007) by incorporating the NIG distribution. Recently, a skewed GARCH-M model based on the so-called z distribution was introduced by Lanne and Saikkonen (2007b) whereas the skewed generalized t distribution of Theodossiou (1998) was proposed by Bali et al. (2008).

Multivariate skewed distributions have been discussed, for instance, by Mencia and Sentana (2004), who consider a multivariate GH distribution in dynamic conditionally heteroskedastic regression models. This distribution nests several asymmetric distributions as well as symmetric t and normal distributions. In Bauwens and Laurent (2005) a multivariate Student distribution is generalized to the asymmetric case and applied to GARCH modeling.

The objective of the second chapter of this thesis is to discuss the importance of allowing for conditional skewness in GARCH models. Therefore, skewed GARCH models are applied to the returns of two real stock index series, and the performance of the models is carefully examined. The fundamental assumption made in the chapter is that the distribution of the innovation η_t can be described by a GH distribution[5]. This distribution contains several well-known special cases of which

---

[4]Related, so-called hyperbolic distributions were introduced to financial econometrics by Eberlein and Keller (1995).

[5]These distributions were initially introduced by Barndorff-Nielsen (1977) for the (logarithm)

the already mentioned NIG distribution is perhaps the best known. Based on earlier literature (e.g. Prause 1999 and Mencia and Sentana 2004) it can be argued that some special case of the GH distribution is capable of modeling the innovation. As the number of available special cases is hardly unsubstantial, we suggest that the number of subclasses considered can be reduced by estimating a GARCH model with GH innovations and using the results obtained to find an appropriate special case (or cases). This approach will also be emphasized by our empirical results.

As already mentioned, the models proposed are likely to have applications to option pricing under GARCH models. The GARCH option pricing model was initially proposed by Duan (1995). Other studies of GARCH option pricing include Engle and Mustafa (1992), Christoffersen and Jacobs (2004), Christoffersen et al. (2006), Badescu et al. (2008) and Stentoft (2008). The original GARCH option pricing model was derived under the assumption of Gaussian errors. However, the work of Stentoft (2008) and Badescu et al. (2008), for example, provides evidence that allowing for non-normality and especially for both conditional skewness and leptokurtosis is indeed profitable in GARCH option pricing.

1.4 Mixture MEM models and conditional autoregressive range

In the third chapter of the thesis we emphasize the fact that modeling non-negative time series is becoming increasingly important in financial econometrics. As already pointed out, a general model for non-negative time series was introduced by Engle (2002b). The so-called multiplicative error model (MEM) specified as

$$x_t = \mu_t \varepsilon_t,$$

---

of particle size of wind blown sands. Properties of these distributions (or related distributions) were later studied e.g. by Barndorff-Nielsen (1978), Barndorff-Nielsen et al. (1978), Jensen (1981), Barndorff-Nielsen et al. (1982), and Barndorff-Nielsen et al. (1992). It may also be noted here that throughout the thesis we will use terms 'distribution' and 'distributions' somewhat loosely in the sense that both terms may refer to a family of distributions or to a distribution with (unknown) parameters.

where $\varepsilon_t \sim D^+(1, \phi^2)$, that is, ε_t has a non-negative distribution with unit mean and variance ϕ^2. The innovations ε_t are also assumed to be independently and identically distributed. The model is expected to be suitable for realized volatility, range-based volatility measures and trading volume. The model was generalized to a mixture MEM model by Lanne (2006)[6], who applied it to the realized volatility of two foreign exchange rates. This formulation will be introduced formally in Chapter 3. This thesis considers mixture MEM models for modeling of the daily price range[7].

Modeling of the daily price range of asset prices with MEM-type models was initially proposed by Chou (2005), who called his model a conditional autoregressive range (CARR) model. As opposed to return based volatility models such as GARCH models, it was argued by Chou (2005, 561-564) that the strength of the range as a volatility estimator was already established in the early eighties e.g. by Parkinson (1980). In particular, it was shown by Parkinson (1980) that the unbiased range-based estimator of the diffusion coefficient of a driftless Brownian Motion was approximately five times more efficient than the estimator based on the daily squared returns (Brunetti and Lildholdt 2007, 40). When modeling the three most common measures of volatility, absolute returns, daily range, and realized volatility, Engle and Gallo (2006) found evidence that range carries additional information to realized volatility. It was also suggested by Brandt and Diebold (2006) that, unlike in the case of realized volatility, the range is not affected by market microstructure noise. By contrast, Chou et al. (2008) conclude that the range is sensitive to outliers and should be replaced with a robust measure such as the quantile range. The empirical volatility forecasting performance of the CARR model was examined by Chou and Wang (2007), who concluded that the CARR model produces sharper volatility forecasts than the GARCH model.

Academic studies on the specification of the CARR model are to the best of our knowledge somewhat rare. Perhaps surprisingly, several model specifications for the log-range have been proposed. As an example one can mention Brandt and Jones

---

[6]These models have been further generalized by Ahoniemi and Lanne (2008) and De Luca and Gallo (2009), who allowed for time-varying mixture weights.

[7]This variable will also be formally defined in Chapter 3.

(2006), who employ the EGARCH model by Nelson (1991) for the log-range. Several other models for the log-range can be found in Chou et al. (2008). Chou (2006) extends his CARR model to an asymmetric ACARR model. Brunetti and Lildholdt (2007) propose both semi-parametric and parametric fractionally integrated (FI)MEM models for the daily ranges of two exchange rates. Multivariate CARR models have also been proposed. Theoretical contributions for these models can be found in Fernandes et al. (2005) and in Lee and Chin (2008).

As mentioned in subchapter 1.2.1, the use of high frequency data is currently one of the most important research topics regarding volatility. One approach is to model high frequency data based on daily volatility indicators such as realized volatility. A new estimator of the volatility was recently suggested by Brunetti et al. (2007) and Christersen and Podolskij (2007), who replaced squared intra-day returns by the high-low range and obtained realized range. The theoretical properties of these volatility estimators or indicators are still partially unknown. However, both Brunetti et al. (2007) and Christersen and Podolskij (2007) note that the realized range is more efficient than the realized volatility, the difference depending on the sampling frequency. Both authors present bias correction procedures in order to account for the market microstructure frictions.

This thesis contributes in several ways to modeling the conditional autoregressive range. First, we consider a general mixture multiplicative error model and present results that allow both strict and weak stationarity of the model to be expressed in a simple way after the model is represented in a first order random coefficient multivariate autoregressive model. Second, we also find a relatively simple expression for the autocorrelation function of the model. In addition, the inverse gamma distribution will be suggested as a suitable alternative for the conditional distribution of the daily price range. The fit of the models is examined through in-sample and out-of-sample examination.

1.5 Modeling financial volatility measures with a VMEM model based on an asymmetric copula

The third topic of the thesis is vector multiplicative error models (VMEM) introduced by Engle (2002b, 429). The model is defined by

$$x_t = \mu_t \odot \varepsilon_t,$$

where μ_t is a $K \times 1$ vector of conditional means, $\odot$ is the element by element (Hadamard) product, and $\varepsilon_t | \Omega_{t-1} \sim D^+(1, \Sigma)$, that is, ε_t $(K \times 1)$ has a non-negative distribution with unit vector as expectation and a general positive definite variance-covariance matrix Σ. The innovations ε_t are assumed to be independently and identically distributed.

The model was applied by Engle and Gallo (2006) to three volatility indicators, absolute returns, daily high-low range, and realized volatility. In their application to Standard & Poor's 500 data it was found that the specifications retained for each indicator contained lagged values of other indicators. In particular, it was discovered that daily range and returns had explanatory power over realized volatility. Further, the model can be employed e.g. to volumes, trades, durations, and various versions of ultra-high frequency based measures of volatility. In Engle and Gallo (2006), however, the components of the innovations ε_t were assumed to be independent. As discussed by Cipollini et al. (2007, 2) assuming independency obviously makes the estimation procedure inefficient, because the correlation of the error terms is not taken into account. Thus, model selection and ensuing interpretation of the model may be inaccurate.

In order to take account of the aforementioned correlation, Cipollini et al. (2006, 2007) introduced copula[8] functions to link together the marginal probability functions. In their more recent work Cipollini et al. (2008) argue that the choice of

---

[8]It may be mentioned here that the copula function is defined as a cumulative distribution function of a continuous (multivariate) uniform random variable defined on the unit hypercube (see e.g. Cipollini et al. 2006, Appendix B).

the copula is often driven by reasons of convenience[9]. More specifically, the authors mention that so-called elliptical copulas may be appealing because they can be employed in relatively large dimensional applications and they can also accommodate tail dependency. However, it seems obvious that elliptical copulas may not be able to correctly describe the dependence structure of the variables. Cipollini et al. (2008) discuss Archimedean copulas as a possible solution but note that they tend to be less useful when the dimension K increases. In our work we show that the symmetrical copulas employed by Cipollini et al. (2006, 2007) may not be optimal when modeling several volatility indicators, such as absolute returns and daily range, simultaneously. We recognize the drawback of Archimedean copulas not being well suited for large dimensional applications, but as our application only concerns two volatility indicators we pursue this approach. Our application shows that allowing for an asymmetric[10] copula considerably improves the fit of the model. Cipollini et al. (2008) argue that in the analysis it may not be of interest to fully specify the distribution of the innovation ε_t if the main focus is the dynamics of μ_t. By contrast, in our application we show that the specification of the distribution of ε_t affects the estimates obtained and thereby the conclusions based on these estimates. Especially, it is discovered that our volatility forecasts have increased predictive power to the so-called VIX index (differences), as the model employed better describes the dependence structure of the components. This application is inspired by Engle and Gallo (2006), who provide evidence that model-based volatility forecasts have significant explanatory power when modeling the value of a market based volatility measure such as the VIX index.

In the earlier work of Engle and Gallo (2006) and Cipollini et al. (2006, 2007) the marginal distributions were specified as a gamma distribution restricted to have a unit mean. In Cipollini et al. (2007) it is mentioned that other distributions, such as inverse gamma, Weibull, Lognormal and their mixtures, may also be employed. In our application, we argue that different marginals for different components are

---

[9]Obviously these reasons include analytical tractability and computational simplicity.
[10]Discussion on asymmetry in the context of copulas can be found in Chapter 4.

indeed needed. In particular, we find evidence that the inverse gamma distribution[11] correctly describes the conditional distribution of the daily price range whereas the gamma distribution appears to fit absolute returns.

1.6 Conclusion

This thesis discusses modeling financial data from various perspectives. In Chapter 2 traditional GARCH modeling is considered. Chapter 3 involves modeling the daily price range whereas Chapter 4 includes modeling return and price range data simultaneously.

When considering GARCH modeling it is discovered that conditional skewness cannot be comprehensively captured even with highly sophisticated smooth transition specification in the conditional variance[12]. In our applications to US and European data so-called normal inverse Gaussian and normal reciprocal gamma (also known as skewed Student's t distribution) prove to be the most relevant special cases of the GH distributions. In both in-sample and out-of-sample comparisons conditional skewness is found to have an impact on the Value-at-Risk forecasting performance of the model.

Although the daily price range was recognized as an efficient estimator of the volatility relatively early e.g. by Parkinson (1980), modeling this variable has not received extensive attention. A multiplicative error model by Engle (2002b) was applied to the range data by Chou (2005) and generalized to the asymmetric case by Chou (2006). As an extension of the MEM model, we propose the mixture MEM model by Lanne (2006) for the daily price range. The mixture structure allows flexibility in both the conditional distribution and the mean dynamics. We find conditions for the strict and weak stationarity and an expression for the autocorrelation function of the mixture MEM model. In an application to Asian data we

---

[11]The inverse gamma distribution is the distribution of the reciprocal of a variable distributed according to the gamma distribution.

[12]We also allow the conditional mean to be time-varying and find evidence of an in-Mean effect in our data. It may be noted that correct specification of both the conditional mean and variance is important when examining the conditional skewness. This is due to the fact that if the conditional mean or variance of the model is misspecified, testing for conditional skewness is biased.

provide evidence that allowing for a more general mixture specification may be advantageous when modeling the daily price range. It is also noted that the inverse gamma distribution may prove relevant in the daily range modeling as it appears to describe the data more accurately than the usually employed gamma distribution. Different specifications are examined through in-sample fit evaluations and out-of-sample forecasting performance tests.

The vector MEM model was introduced by Engle (2002b). Among others, Cipollini et al. (2006, 2007) applied the model to several volatility indicators and used copula functions to describe the dependence structure of the error terms. In the fourth chapter of this thesis we show that symmetric copulas proposed by Cipollini et al. (2006, 2007) may not be optimal for the variables considered if the dependence of the errors is markedly asymmetric. By employing a proper asymmetric copula we show that the fit of the model used for two volatility indicators of the S&P 500 data can be considerably improved. Different marginal distributions are proposed for each variable. We show that model based forecasts derived from our asymmetric VMEM model outperform those of the symmetric model when forecasts are used to explain the behavior of a market-based volatility measure, the VIX index, both in-sample and out-of-sample. We also verify the finding of Engle and Gallo (2006) that the daily range has explanatory power over other volatility indicators.

References

Aas, K. & I. H. Haff 2006: The generalized hyperbolic skew Student's t-distribution, *Journal of Financial Econometrics*, 4, 275-309.

Ahoniemi, K. & M. Lanne 2008: Implied volatility with time-varying regime probabilities, HECER Discussion Paper No 246.

Alexander, C. & E. Lazar 2006: Normal mixture GARCH(1,1): Applications to exchange rate modelling, *Journal of Applied Econometrics*, 21, 307-336.

Andersen, T. G., T. Bollerslev, F. X. Diebold & H. Ebens 2001a: The distribution of realized stock return volatility, *Journal of Financial Economics*, 61, 43-76.

Andersen, T. G., T. Bollerslev, F. X. Diebold & P. Labys 2001b: The distribution of realized exchange rate volatility, *Journal of the American Statistical Association*, 96, 42-55.

Andersen, T. G., T. Bollerslev, F. X. Diebold & P. Labys 2003: Modeling and forecasting realized volatility, *Econometrica*, 71, 579-625.

Andersson, J. 2001: On the normal inverse Gaussian stochastic volatility model, *Journal of Business & Economic Statistics*, 19, 44-54.

Bachelier, L. 1900: Theory of speculation (translation of 1900 French edition), in P. H. Cootner, ed., The Random Character of Stock Market Prices, Cambridge: M.I.T. Press, 1964, 17-78.

Badescu, A. & R. Kulperger 2008: GARCH option pricing: A semiparametric approach, *Insurance: Mathematics & Economics*, 43, 69-84.

Badescu, A., R. Kulperger & E. Lazar 2008: Option valuation with normal mixture GARCH models, *Studies in Nonlinear Dynamics & Econometrics*, 12, Issue 2, Article 5.

Bali, T. G., H. Mo & Y. Tang 2008: The role of autoregressive conditional skewness and kurtosis in the estimation of conditional VaR, *Journal of Banking & Finance*, 32, 269-282.

Barndorff-Nielsen, O. 1977: Exponentially decreasing distributions for the logarithm of particle size, *Proceedings of the Royal Society of London A*, 353, 401-419.

Barndorff-Nielsen, O. 1978: Hyperbolic distributions and distributions on hyperbolae, *Scandinavian Journal of Statistics*, 5, 151-157.

Barndorff-Nielsen, O. E. 1997: Normal inverse Gaussian distributions and stochastic volatility modeling, *Scandinavian Journal of Statistics*, 24, 1-13.

Barndorff-Nielsen, O., P. Blæsild & C. Halgreen 1978: First hitting time models for the generalized inverse Gaussian distribution, *Stochastic Processes and Their Applications*, 7, 49-54.

Barndorff-Nielsen, O. E., P. Blæsild & V. Seshadri 1992: Multivariate distributions with generalized inverse Gaussian marginals, and associated Poisson mixtures, *The Canadian Journal of Statistics*, 20, 109-120.

Barndorff-Nielsen, O. E., J. Kent & M. Sørensen 1982: Normal variance-mean mixtures and z distributions, *International Statistical Review*, 50, 145-159.

Bauwens, L. & S. Laurent 2005: A new class of multivariate skew densities, with application to GARCH models, *Journal of Business & Economic Statistics*, 23, 346-354.

Bauwens, L., S. Laurent & J. V. K. Rombouts 2006: Multivariate GARCH models: A survey, *Journal of Applied Econometrics*, 21, 79-109.

Bera, A. K. & M. L. Higgins 1993: ARCH models: Properties, estimation and testing, *Journal of Economic Surveys*, 7, 305-366.

Black, F. & M. Scholes 1973: The pricing of options and corporate liabilities, *Journal of Political Economy*, 81, 637-654.

Bollerslev, T. 1986: Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 31, 301-327.

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

Bollerslev, T. 1990: Modeling the coherence in short-run nominal excess rates: A multivariate generalized ARCH model, *The Review of Economics and Statistics*, 72, 498-505.

Bollerslev, T. 2001: Financial econometrics: Past developments and future challenges, *Journal of Econometrics*, 100, 41-51.

Bollerslev, T. 2008: Glossary to ARCH (GARCH), CREATES Research Paper 2008-49.

Bollerslev, T., R. Y. Chou & K. F. Kroner 1992: ARCH modeling in finance: A review of the theory and empirical evidence, *Journal of Econometrics*, 52, 5-59.

Bollerslev, T., R. F. Engle & D. B. Nelson 1994: ARCH models, in R. F. Engle & D. L. McFadden, ed., Handbook of Econometrics IV, Amsterdam: Elsevier Science, 2961-3038.

Bollerslev, T., R. F. Engle & J. Wooldridge 1988: A capital asset pricing model with time varying covariances, *Journal of Political Economy*, 96, 116-131.

Brandt, M. & F. X. Diebold 2006: A no-arbitrage approach to range-based estimation of return covariances and correlations, *Journal of Business*, 79, 61-74.

Brandt, M. & C. Jones 2006: Volatility forecasting with range-based EGARCH models, *Journal of Business & Economic Statistics*, 24, 470-486.

Brunetti, C. & P. M. Lildholdt 2007: Time series modelling of daily log-price ranges for SF/USD and USD/GBP, *The Journal of Derivatives*, 15, 2, 39-59.

Campbell, J. Y. & L. Hentschel 1992: No news is good news: An asymmetric model of changing volatility in stock returns, *Journal of Financial Economics*, 31, 281-318.

Chorro, C., D. Guégan & F. Ielpo 2008a: Option pricing under GARCH models with Generalized Hyperbolic innovations (I): Methodology, CES Working Papers 2008.37.

Chorro, C., D. Guégan & F. Ielpo 2008b: Option pricing under GARCH models with Generalized Hyperbolic innovations (II): Data and results, CES Working Papers 2008.47.

Chou, H.-C. & D. Wang 2007: Forecasting volatility on the U.K. stock market: A test of the conditional autoregressive range model, *International Research Journal of Finance and Economics*, 10, 7-13.

Chou, R. Y. 2005: Forecasting financial volatilities with extreme values: The conditional autoregressive range (CARR) model, *Journal of Money, Credit and Banking*, 37, 561-582.

Chou, R. Y. 2006: Modeling the asymmetry of stock movements using price ranges, *Advances in Econometrics*, 20A, 231-257.

Chou, R. Y., H. Chou & N. Liu 2008: Range volatility models and their applications in finance, in C.-F. Lee & A. C. Lee, ed., Handbook of Quantitative Finance and Risk Management (forthcoming).

Christensen, K. & M. Podolskij 2007: Realized range-based estimation of integrated variance, *Journal of Econometrics*, 141, 323-349.

Christie, A. A. 1982: The stochastic behavior of common stock variances: Value, leverage, and interest rate effects, *Journal of Financial Economics*, 10, 402-432.

Christoffersen, P., S. Heston & K. Jacobs 2006: Option valuation with conditional skewness, *Journal of Econometrics*, 131, 253-284.

Christoffersen, P. & K. Jacobs 2004: Which GARCH model for option valuation?, *Management Science*, 50, 1204-1221.

Cipollini, F., R. F. Engle & G. M. Gallo 2006: Vector multiplicative error models: Representation and inference, NBER Working Paper Series, 12690.

Cipollini, F., R. F. Engle & G. M. Gallo 2007: A model for multivariate non-negative valued processes in financial econometrics, Working Papers (2007/16), DSUNIFI, Dipartimento di Statistica "Giuseppe Parenti".

Cipollini, F., R. F. Engle & G. M. Gallo 2008: Semiparametric vector MEM, Working paper, SSRN.

De Luca, G. & G. M. Gallo 2009: Time-varying mixture weights in mixture autoregressive conditional duration models, *Econometric Reviews*, 28, 102-120.

Degiannakis, S. & E. Xekalaki 2004: Autoregressive conditional heteroskedasticity (ARCH) models: A review, *Quality Technology & Quantitative Management*, 1, 271-324.

Diebold, F. X. 2004: The Nobel memorial prize for Robert F. Engle, *Scandinavian Journal of Economics*, 106, 165-185.

Duan, J.-C. 1995: The GARCH option pricing model, *Mathematical Finance*, 5, 13-32.

Duan, J.-C., P. Ritchken & Z. Sun 2006: Approximating GARCH-jump models, jump-diffusion processes, and option pricing, *Mathematical Finance*, 16, 21-52.

Eberlein, E. & U. Keller 1995: Hyperbolic distributions in finance, *Bernoulli*, 1, 281-299.

Elliot, R. J., T. K. Siu & L. Chan 2006: Option pricing for GARCH models with Markov switching, *International Journal of Theoretical and Applied Finance*, 9, 825-841.

Engle, R. F. 1982: Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, 50, 987-1007.

Engle, R. F. 2000: The econometrics of ultra-high-frequency data, *Econometrica*, 68, 1-22.

Engle, R. F. 2001: Financial econometrics - A new discipline with new methods, *Journal of Econometrics*, 100, 53-56.

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. 2004: Risk and volatility: Econometric models and financial practice, *American Economic Review*, 94, 405-420.

Engle, R. F. & G. M. Gallo 2006: A multiple indicators model for volatility using intra-daily data, *Journal of Econometrics*, 131, 3-27.

Engle, R. F., G. M. Gallo & M. Velucci 2008: A MEM-based analysis on volatility spillovers in East Asian financial markets, Working Papers (2008/09), DSUNIFI, Dipartimento di Statistica "Giuseppe Parenti".

Engle, R. F. & K. Kroner 1995: Multivariate simultaneous GARCH, *Econometric Theory*, 11, 122-150.

Engle, R. F., D. M. Lilien & R. P. Robins 1987: Estimating time varying risk premia in the term structure: The ARCH-M model, *Econometrica*, 55, 391-407.

Engle, R. F. & C. Mustafa 1992: Implied ARCH models from options prices, *Journal of Econometrics*, 52, 289-311.

Engle, R. F., V. Ng & M. Rothschild 1990: Asset pricing with a factor ARCH covariance structure: Empirical estimates for treasury bills, *Journal of Econometrics*, 45, 213-238.

Engle, R. F. & J. E. Russel 1998: Autoregressive conditional duration: A new model for irregularly spaced transaction data, *Econometrica*, 66, 1127-1162.

Fernandes, M., B. de Sá Mota & G. Rocha 2005, A multivariate conditional autoregressive range model, *Economics Letters*, 86, 435-440.

Forsberg, L. & T. Bollerslev 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.

Glosten, L. R., R. Jagannathan & D. E. Runkle 1993: On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance*, 48, 1779-1801.

Haas, M., S. Mittnik & M. S. Paolella 2004: Mixed normal conditional heteroskedasticity, *Journal of Financial Econometrics*, 2, 211-250.

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

Jensen, J. J. 1981: On the Hyperboloid distribution, *Scandinavian Journal of Statistics*, 8, 193-206.

Jensen, M. B. & A. Lunde 2001: The NIG-S&ARCH model: A fat-tailed, stochastic, and autoregressive conditional heteroskedastic volatility model, *Econometrics Journal*, 4, 319-342.

Jondeau, E. & M. Rockinger 2006: The copula-GARCH model of conditional dependencies: An international stock market application, *Journal of International Money and Finance*, 25, 827-853.

Kiliç, R. 2007: Conditional volatility and distribution of exchange rates: GARCH and FIGARCH models with NIG distribution, *Studies in Nonlinear Dynamics & Econometrics*, 11, Issue 3, Article 1.

Lanne, M. 2006: A mixture multiplicative error model for realized volatility, *Journal of Financial Econometrics*, 4, 594-616.

Lanne, M. & P. Saikkonen 2007a: A multivariate generalized orthogonal factor GARCH model, *Journal of Business & Economic Statistics*, 25, 61-75.

Lanne, M. & P. Saikkonen 2007b: Modeling conditional skewness in stock returns, *European Journal of Finance*, 13, 691-704.

Lee, O. & D. W. Shin 2008: Geometric ergodicity and β-mixing property for a multivariate CARR model, *Economics Letters*, 100, 111-114.

Manganelli, S. 2005: Duration, volume and volatility impact of trades, *Journal of Financial Markets*, 8, 377-399.

Martens, M. & D. van Dijk 2007: Measuring volatility with the realized range, *Journal of Econometrics*, 138, 181-207.

Mencia, F. J. & E. Sentana 2004: Estimation and testing of dynamic models with generalized hyperbolic innovations, CEMFI Working Paper, 0411.

Merton, R. C. 1973: An intertemporal capital asset pricing model, *Econometrica*, 41, 867-887.

Merton, R. C. 1973: Theory of rational option pricing, *Bell Journal of Economics and Management Science*, 4, 141-183.

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

Palm, F. C. 1996: GARCH models of volatility, in G. S. Maddala & C. R. Rao, ed., Handbook of Statistics 14, Amsterdam: Elsevier Science, 209-240.

Parkinson, M. 1980: The extreme value method for estimating the variance of the rate of return, *Journal of Business*, 53, 61-65.

Patton, A. J. 2006: Modelling asymmetric exchange rate dependence, *International Economic Review*, 47, 527-556.

Prause, K. 1999: The generalized hyperbolic model: Estimation, financial derivatives, and risk measures. Dissertation zur Erlangung des Doktograges der Mathematischen Fakultät der Albert-Ludwigs-Universität Freiburg i. Br.

Silvennoinen, A. & T. Teräsvirta 2008: Multivariate GARCH models, CREATES Research Paper 2008-6.

Siu, T. K., H. Tong & H. Yang 2004: On pricing derivatives under GARCH models: A dynamic Gerper-Shiu approach, *North American Actuarial Journal*, 8, 17-32.

Stentoft, L. 2008: American option pricing using GARCH models and the normal inverse Gaussian distribution, CREATES Research Paper 2008-41.

Theodossiou, P. 1998: Financial data and the skewed generalized t distribution, *Management Science*, 44, 1650-1661.

2 GARCH modeling with generalized hyperbolic distributions

The outline of this chapter is as follows. The GH distribution is defined in sub-chapter 2.1 and, its special cases, the NIG distribution and the NRG, that prove most relevant in our applications, are discussed in some detail. In subchapter 2.2, GARCH-M models based on these distributions are considered along with various specifications for the conditional variance. A brief discussion on parameter estimation is also included. Empirical applications to returns of Standard & Poor's 500 and Amsterdam EOE indices are presented in subchapter 2.3. The usefulness of allowing for conditional skewness is illustrated by Value-at-Risk applications using both in-sample and out-of-sample analysis. Conclusions are drawn in subchapter 2.4.

2.1 Generalized hyperbolic distributions

The density function of the GH distribution expressed with the parameterization most suitable for our purposes[13] is (e.g. Prause 1999, Appendix C)

$$
\begin{aligned}
f(x; \lambda, \alpha, \beta, \delta, \mu) \;=\; & a(\lambda, \alpha, \beta, \delta) \left(1 + \left(\frac{x-\mu}{\delta} \right)^2 \right)^{(\lambda - 1/2)/2} \\
& \times K_{\lambda - 1/2} \left(\alpha \sqrt{1 + \left(\frac{x-\mu}{\delta} \right)^2} \right) \exp \left(\beta \left(\frac{x-\mu}{\delta} \right) \right), \quad (2)
\end{aligned}
$$

$x \in \mathbb{R}$, where

$$
a(\lambda, \alpha, \beta, \delta) = \frac{(\alpha^2 - \beta^2)^{\lambda/2}}{\sqrt{2\pi} \alpha^{\lambda - 1/2} \delta K_\lambda \left(\sqrt{\alpha^2 - \beta^2} \right)}
$$

---

[13]For other parameterizations of the GH distribution, see e.g. Prause (1999), Eberlein and v. Hammerstein (2003) or Aas and Haff (2006).

and K_λ is the modified Bessel function of the third kind.[14] The parameters are interpreted as follows: $\mu \in \mathbb{R}$ is a location parameter and $\delta > 0$ is a scale parameter. The parameter $0 \leq |\beta| < \alpha$ describes the skewness and $\alpha > 0$ gives the kurtosis. If $\beta = 0$, the distribution is symmetric. The parameter $\lambda \in \mathbb{R}$ characterizes certain subclasses of the distribution and considerably influences the size of the probability mass contained in the tails of the distribution (Eberlein and v. Hammerstein 2003).

If the random variable x has a GH distribution, we write $x \sim GH(\lambda, \alpha, \beta, \delta, \mu)$. This distribution belongs in a location-scale family of distributions[15], meaning that

$$x \sim GH(\lambda, \alpha, \beta, \delta, \mu) \Leftrightarrow \frac{x - \mu}{\delta} \sim GH(\lambda, \alpha, \beta, 1, 0).$$

As will be seen in subchapter 2.3, this property is convenient in GARCH applications. The moment-generating function of the GH distribution can be found in Eberlein (2001, Equation 3.5).

The mean and variance of the GH distribution are given by

$$Ex = \mu + \delta \frac{\beta}{\gamma} \frac{K_{\lambda+1}(\gamma)}{K_\lambda(\gamma)} \tag{3}$$

and

$$Var(x) = \delta^2 \left(\frac{K_{\lambda+1}(\gamma)}{\gamma K_\lambda(\gamma)} + \frac{\beta^2}{\alpha^2 - \beta^2} \left[\frac{K_{\lambda+2}(\gamma)}{K_\lambda(\gamma)} - \left(\frac{K_{\lambda+1}(\gamma)}{K_\lambda(\gamma)} \right)^2 \right] \right), \tag{4}$$

where $\gamma = \sqrt{\alpha^2 - \beta^2} > 0$ (Prause 1999, Appendix C).

We shall next discuss two special cases. For more special cases of the GH distribution and their properties, see for example Prause (1999) or Eberlein and v. Hammerstein (2003).

The first special case of the GH distribution we discuss is the NIG distribution.

---

[14]For more information on the modified Bessel function see, for example, the Appendix in Jørgensen (1982) or Appendix B in Prause (1999).

[15]In general, a random variable y belongs to a location-scale family of distributions if the cumulative distribution function of y can be expressed as $F(y; \mu, \sigma) = \Phi \left(\frac{y-\mu}{\sigma} \right)$, where $-\infty < \mu < \infty$ is a location parameter, $\sigma > 0$ is a scale parameter and Φ is the cumulative distribution function of y when $\mu = 0$ and $\sigma = 1$.

The density function of the NIG distribution is obtained by assuming that $\lambda = -1/2$ in Equation (2). Because the modified Bessel function satisfies $K_{-1/2}(x) = K_{1/2}(x) = (\pi/2x)^{-1/2} e^{-x}$, we obtain

$$f(x; \alpha, \beta, \delta, \mu) = \frac{\alpha}{\pi\delta} \exp\left[\sqrt{\alpha^2 - \beta^2} + \beta\left(\frac{x-\mu}{\delta}\right)\right] \frac{K_1\left(\alpha\sqrt{1 + \left(\frac{x-\mu}{\delta}\right)^2}\right)}{\sqrt{1 + \left(\frac{x-\mu}{\delta}\right)^2}}, \quad (5)$$

where $x, \mu \in \mathbb{R}$, $\delta > 0$ and $0 \leq |\beta| < \alpha$ (Barndorff-Nielsen 1997, Equation 2.4). The parameters are interpreted in the same way as in the case of the GH distribution. We write $x \sim NIG(\alpha, \beta, \delta, \mu)$ to signify that the random variable x has an NIG distribution.

Well-known expressions (see e.g. Jensen and Lunde 2001, 325) for the first and second cumulants of an NIG distributed random variable are

$$\varkappa_1 = \mu + \frac{\rho\delta}{\sqrt{1 - \rho^2}} = \mu + \frac{\beta}{\gamma}\delta = Ex$$

and

$$\varkappa_2 = \frac{\delta^2}{\alpha\left(\sqrt{1 - \rho^2}\right)^3} = \frac{\alpha^2}{\gamma^3}\delta^2 = Var(x),$$

where $\gamma = \sqrt{\alpha^2 - \beta^2}$ and $\rho = \beta/\alpha$. Straightforwardly from the expressions of the third and fourth cumulant (Jensen and Lunde 2001, 325) we obtain the skewness

$$Skw(x) = \frac{\varkappa_3}{(\varkappa_2)^{3/2}} = 3\frac{\rho}{\sqrt[4]{1 - \rho^2}\sqrt{\alpha}},$$

and the kurtosis

$$Kts(x) = \frac{\varkappa_4}{(\varkappa_2)^2} + 3 = 3\left(1 + \frac{4\rho^2 + 1}{\alpha\sqrt{1 - \rho^2}}\right).$$

The second special case of the GH distribution discussed in detail in this subchapter is the so-called NRG distribution. The NRG distribution has the density

function

$$f(x; \lambda, \beta, \delta, \mu) = \left(\frac{2}{\pi}\right)^{1/2} \frac{\exp[\beta\left(\frac{x-\mu}{\delta}\right))]}{\delta 2^{-\lambda}\Gamma(-\lambda)} \frac{\left(\sqrt{1 + \left(\frac{x-\mu}{\delta}\right)^2}\right)^{(\lambda-1/2)}}{|\beta|^{(\lambda-1/2)}}$$

$$\times K_{\lambda-1/2}\left(|\beta|\sqrt{1 + \left(\frac{x-\mu}{\delta}\right)^2}\right), \tag{6}$$

where $x, \mu \in \mathbb{R}$, $\lambda < 0$, and $\beta, \delta > 0$. A positive (negative) sign of β in the exponential function implies positive (negative) skewness. Other parameters are interpreted as in the case of the GH distribution. If the random variable x has an NRG distribution, we write $x \sim NRG(\lambda, \beta, \delta, \mu)$.

The density function of the NRG distribution is obtained as a limit from (2) when $\alpha \to |\beta| > 0$. This is seen by using the well-known asymptotic property of the modified Bessel function that, for $x \to 0$,

$$K_\lambda(x) \sim \frac{1}{2}\Gamma(-\lambda)\left(\frac{x}{2}\right)^\lambda, \ \lambda < 0$$

(Abramowitz and Stegun 1970, Equation 6.3.8; Eberlein and v. Hammerstein 2003, Equation 3.2). Thus, we have

$$\frac{1}{\delta^\lambda K_\lambda(\sqrt{\alpha^2 - \beta^2})} \sim \frac{2^{\lambda+1}\delta^{-\lambda}}{(\alpha^2 - \beta^2)^{\lambda/2}\Gamma(-\lambda)}, \tag{7}$$

and, using this in (2), gives (6).

The Student's t distribution is obtained as a special case of the NRG distribution by letting $\beta \to 0$ in Equation (6). We obtain

$$f(x; \lambda, \delta, \mu) = \frac{\Gamma(-\lambda + 1/2)}{\Gamma(-\lambda)\sqrt{\pi\delta^2}}\left(1 + \frac{(x-\mu)^2}{\delta^2}\right)^{\lambda-1/2},$$

the density of a scaled and shifted t distribution. Specializing to $\delta^2 = -2\lambda$ gives the density of a shifted t distribution with the degrees of freedom parameter equal to -2λ, where $\lambda < -1$.

By using (7) in (3) and (4) we obtain for the NRG distribution

$$Ex = \mu + \delta \frac{\beta}{2(-\lambda - 1)},$$

and

$$Var(x) = \delta^2 \left[\frac{1}{2(-\lambda - 1)} + \frac{\beta^2}{4(-\lambda - 1)^2(-\lambda - 2)} \right],$$

where $\lambda < -2$.

The NRG distribution does not have a moment generating function. However, the moments can be derived after noting that the distribution has a normal variance-mean mixture representation.

According to Barndorff-Nielsen et al. (1982, 145) the distribution of a random variable x is a normal variance-mean mixture with location μ, drift $\bar{\beta} = \beta/\delta$, and non-negative mixing variable y if, for a given y, the distribution of x is normal with mean $\mu + \bar{\beta}y$ and variance y. The distribution of y is referred to as a mixing distribution. The same distribution can alternatively be obtained by assuming the representation

$$x = \mu + \bar{\beta}y + y^{1/2}\epsilon, \tag{8}$$

where $\epsilon \sim n.i.d.(0, 1)$ is independent of y. If $\bar{\beta} = 0$, the distribution is symmetric and called normal variance mixture. The (unconditional) distribution of x is determined by specifying the distribution of y.

The NRG distribution is obtained by assuming that the mixing variable in (8) has a *reciprocal* or *inverse gamma* distribution $RG(-\lambda, \delta^2/2)$. The raw moments of the reciprocal gamma distribution are given by (Jørgensen 1982, 13)

$$E(y^i) = \frac{\delta^{2i}\Gamma(-\lambda - i)}{2^i\Gamma(-\lambda)}, \tag{9}$$

where the i-th moment exists if $i < -\lambda$. The skewness and kurtosis of the NRG distribution are obtained by using (8) with the well known general expressions of

skewness and kurtosis. The results are

$$Skw(x) = \frac{-3\bar{\beta}(Ey)^2 + 2\bar{\beta}^3(Ey)^3 - 3\bar{\beta}^3 EyEy^2 + 3\bar{\beta}Ey^2 + \bar{\beta}^3 Ey^3}{[Ey + \bar{\beta}^2(E(y^2) - (Ey)^2)]^{3/2}}$$

and

$$
\begin{aligned}
Kts(x) &= \frac{6\bar{\beta}^2(Ey)^3 - 3\bar{\beta}^4(Ey)^4 + 6\bar{\beta}^4(Ey)^2 E(y^2)}{[Ey + \bar{\beta}^2(E(y^2) - (Ey)^2)]^2} \\
&+ \frac{3E(y^2) + 6\bar{\beta}^2 E(y^3) + \bar{\beta}^4 E(y^4)}{[Ey + \bar{\beta}^2(E(y^2) - (Ey)^2)]^2},
\end{aligned}
$$

where $E(y^i)$, $i = 1, 2, 3, 4$, are obtained from (9).

Other special cases of the GH distribution are obtained, for example, by choosing $\lambda \in 1/2\,\mathbb{Z}$ ($\lambda \neq -1/2$), or by letting $\delta \to 0$. The decide $\lambda = 1$ leads to the so-called hyperbolic distribution (see e.g. Prause 1999) whereas when $\delta \to 0$ we get the variance-gamma distribution of Madan et al. (1998).

To get an idea of the general shape of the distributions discussed graphs of the logarithmic densities are presented in Figure 1. Having the applications to GARCH modeling in mind, the mean and variance are set at zero and unity respectively. The shape parameters of the skewed distributions with smallest skewness have been chosen so that the skewness and excess kurtosis equal $-.25$ and 1 respectively. These choices are in line with the empirical findings in subchapter 2.4.

The figure on the left presents the NIG distribution and standardized t distribution with unit excess kurtosis. Logarithmic densities make visible differences in the tails and show the effect of the skewness of the NIG distribution. Different shapes of the tails can be observed when the NRG distribution is plotted against the t distribution. In particular, the logarithmic density reveals the heavy left tail of the NRG distribution. The figures also demonstrate changes in the densities as the values of the skewness parameters increase.
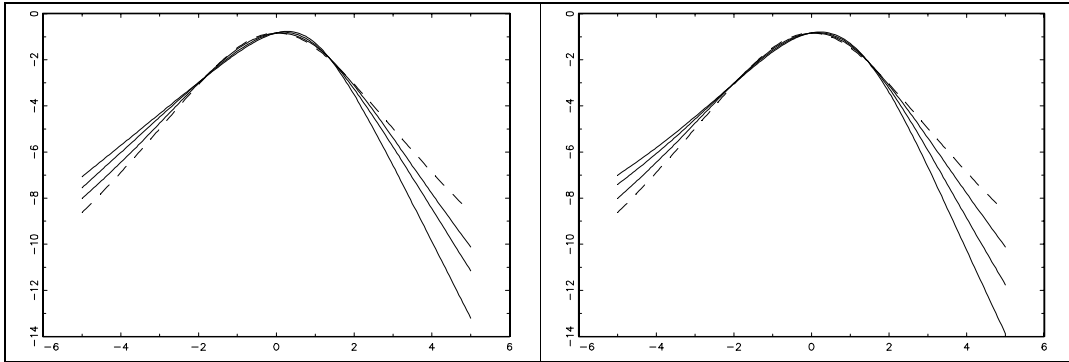
Figure 1. Logarithmic densities of the NIG (left), NRG (right) and t distributions (dashed lines).

2.2 GARCH-M models

This subchapter discusses the GARCH-in-Mean models used in empirical applications of the chapter. First the model given in Equation (1) is specified more precisely. We wish to allow for both a skewed conditional distribution and an asymmetric specification for the conditional variance. Thus, we consider a general non-linear GARCH model which contains as special cases the ST-GARCH model by Lanne and Saikkonen (2005) and the asymmetric GJR-GARCH model by Glosten et al. (1993) as well as an extension of the GARCH model proposed in Lanne and Saikkonen (2007). The normal distribution and the t distribution are considered as conventional symmetric distributions for the error term η_t in (1), whereas the asymmetric distributions presented in the previous subchapter are their asymmetric alternatives. At the end, the estimation of the parameters of the models is discussed.

2.2.1 Specification of the conditional mean and variance

A commonly used specification for the conditional mean in Equation (1), also adopted here, is $m_t = \phi_0 + \phi_1 r_{t-1} + ... + \phi_l r_{t-l} + \nu h_t$. With this specification the GARCH-M model considered reads as

$$r_t = \phi_0 + \phi_1 r_{t-1} + ... + \phi_l r_{t-l} + \nu h_t + h_t^{1/2}\eta_t, \tag{10}$$

where $\phi_1, ..., \phi_l$ and ν are real valued parameters and h_t and η_t are as in (1). For stationarity, the roots of the polynomial $1 - \phi_1 z - ... - \phi_l z^l$ are required to lie outside the unit circle. Any available model can be used to model conditional heteroskedasticity.

The specification we consider is given by

$$h_t = a_0 + d_1 G_1(h_{t-1}) + \sum_{j=1}^{p} b_j h_{t-j} + \sum_{j=1}^{q} a_j u_{t-j}^2 + \sum_{j=1}^{q} c_j I_{t-j} u_{t-j}^2, \qquad (11)$$

where

$$u_t = r_t - m_t - \kappa h_t^{1/2} \qquad (12)$$

with κ a real valued parameter, and

$$I_{t-j} = \begin{cases} 0, & \text{if } u_{t-1} \geq 0, \\ 1, & \text{if } u_{t-1} < 0. \end{cases}$$

Clearly, a model encompassing the standard GJR-GARCH and ST-GARCH specification is obtained if $\kappa = 0$. The reason for allowing other possibilities is that in the case of asymmetric conditional distributions it may not be clear what is the best way to center the series (cf. Lanne and Saikkonen 2007, 696). If $d_1 = 0$ we have the GJR-GARCH by Glosten et al. (1993), which assumes that negative errors contribute more to the conditional variance. When $c_j = 0$, $j = 1, ..., q$, the model corresponds the ST-GARCH model by Lanne and Saikkonen (2005).

For non-negativity of h_t, the parameters in (11) are supposed to satisfy $a_0 > 0$, $b_j \geq 0$, $a_j \geq 0$, and $d_1 \geq 0$ whereas $G_1: (0, \infty) \to [0, 1]$ is an increasing function which depends on the parameters. The function G_1 can be used to allow for a smooth shift in the level parameter a_0. In particular, when h_{t-1} takes small values the process (11) is close to a GJR-GARCH(p, q) process with level parameter a_0. As the value of h_{t-1} increases, the process approaches a GJR-GARCH(p, q) process with level parameter $a_0 + d_1$. In our empirical applications we follow Lanne and Saikkonen (2005) and choose the function G_1 as the cumulative distribution function

of a standard gamma distribution. Specifically, we assume

$$G_1(h_{t-1}) = \int_0^{h_{t-1}} \frac{x^{g_1-1}}{\Gamma(g_1)} \exp(-x) dx,$$

where $g_1 > 0$ is the parameter of the distribution.

As discussed by Lanne and Saikkonen (2005), a major motivation of the nonlinear function G_1 in (11) is that in many applications of the standard GARCH(1,1) model the sum of the parameters a_1 and b_1 is estimated relatively close to unity. Hence, the stationarity condition $a_1 + b_1 < 1$ is nearly violated. This can lead to relatively poor volatility forecasts as the model exaggerates the persistence of the volatility. In their empirical example the authors demonstrate that a ST-GARCH model of the form (11) may then be useful.

The model specification is completed by specifying the distribution of the error term η_t. One may simply assume that $\eta_t \sim N(0,1)$, or, as in Bollerslev (1987), that $\eta_t \sim t\left(\lambda, \sqrt{2(-\lambda-1)}, 0\right)$, that is, a t distribution standardized to have unit variance and with degrees of freedom -2λ, $\lambda < -1$. In subchapter 2.4, the latter will be used as benchmark to which the skewed distributions will be compared.

When $\kappa \neq 0$, the usual stationarity conditions of the GARCH process are not directly applicable. In the case where the extension is applied with the ST-GARCH(1,1) and/or GJR-GARCH(1,1), the stationarity conditions can be concluded from Theorems 1 and 2 of Meitz and Saikkonen (2008). In the important special case $p = q = 1$ and $c_1 = d_1 = 0$ Lanne and Saikkonen (2007) noted that a sufficient condition for stationarity can be obtained from Corollary 6 of Carrasco and Chen (2002). Specifically, under mild conditions on the distribution of η_t (see Carrasco and Chen 2002, 23) it suffices to assume that

$$E(b_1 + a_1(\eta_t - \kappa)^2)^k < 1, \ k \geq 1,$$

where k is an integer. Then $[\ h_t \quad u_t\ ]'$ can be treated as a stationary process and the same is true for r_t. Moreover, $Eh_t^k < \infty$, $E|u_t|^{2k} < \infty$.

2.2.2 GARCH-M models based on GH distributions

In this subchapter it is assumed that the error term η_t has a GH distribution. As already mentioned, the GH distribution itself is only used for finding an eligible sub-class to be applied. Obtaining GARCH-M-models based on these subclasses is discussed after the formulation of the GARCH-M-GH model.

First assume that the innovation η_t has a GH distribution. The distribution is standardized to have zero mean and unit variance by constraining the scale and location parameters as

$$\delta(\lambda, \alpha, \beta) = \left(\frac{K_{\lambda+1}(\gamma)}{\gamma K_\lambda(\gamma)} + \frac{\beta^2}{\alpha^2 - \beta^2} \left[\frac{K_{\lambda+2}(\gamma)}{K_\lambda(\gamma)} - \left(\frac{K_{\lambda+1}(\gamma)}{K_\lambda(\gamma)} \right)^2 \right] \right)^{-1/2}$$

and

$$\mu(\lambda, \alpha, \beta) = -\frac{\delta(\lambda, \alpha, \beta)\beta}{\gamma} \frac{K_{\lambda+1}(\gamma)}{K_\lambda(\gamma)},$$

where $\gamma = \sqrt{\alpha^2 - \beta^2}$. The assumed distribution of the innovation is now given as

$$\eta_t \sim GH\left(\lambda, \alpha, \beta, \delta(\lambda, \alpha, \beta), \mu(\lambda, \alpha, \beta)\right), \tag{13}$$

so that the GARCH-M model is defined by (10), (11), and (13). The implied conditional distribution of r_t can then be expressed as

$$r_t | \Omega_{t-1} \sim GH\left(\alpha, \beta, h_t^{1/2}\delta(\lambda, \alpha, \beta), m_t - h_t^{1/2}\mu(\lambda, \alpha, \beta)\right).$$

Applying Equations (3) and (4) it is verified that m_t and h_t are the conditional mean and variance of r_t given the information set $\Omega_{t-1} = \{r_{t-1}, r_{t-2}, \dots\}$.

GARCH models with NIG and NRG innovations are now obtained simply by using $\lambda = -1/2$ and $\alpha \to |\beta|$ in the preceding formulae. In the NIG case we have

$$\eta_t \sim NIG\left(\alpha, \beta, \gamma^{3/2}/\alpha, -\gamma^{1/2}\beta/\alpha\right), \tag{14}$$

and the model is defined by (10), (11), and (14). Obviously, the conditional distribution of r_t is given as

$$r_t|\Omega_{t-1} \sim NIG\left(\alpha, \beta, h_t^{1/2}\gamma^{3/2}/\alpha, m_t - h_t^{1/2}\gamma^{1/2}\beta/\alpha\right). \tag{15}$$

The conditional density of r_t given Ω_{t-1} is obtained from this and Equation (5).

In the NGR case

$$\eta_t \sim NRG(\lambda, \beta, \delta(\lambda, \beta), -\beta\delta(\lambda, \beta)/\left[2(-\lambda - 1)\right]) \tag{16}$$

where $\delta(\lambda, \beta) = \left[\frac{1}{2(-\lambda-1)} + \frac{\beta^2}{4(-\lambda-1)^2(-\lambda-2)}\right]^{-1/2}$, $\lambda < -2$. Now the model is defined by (10), (11), and (16). With the notation from the previous chapters the conditional distribution of r_t can be expressed as

$$r_t|\Omega_{t-1} \sim NRG(\lambda, \beta, h_t^{1/2}\delta(\lambda, \beta), m_t - h_t^{1/2}\beta\delta(\lambda, \beta)/\left[2(-\lambda - 1)\right]), \tag{17}$$

where $\delta(\lambda, \beta)$ is as above and $\lambda < -2$. From (6) and (17) we find the conditional density of r_t.

If needed, other special cases of GARCH-M-GH models can be derived in the same way as long as the resulting distributions can be represented with the parameters applied here. This is not the case for the so-called variance-gamma distribution of Madan et al. (1998), because this distribution is obtained when $\delta \to 0$.[16]

2.2.3 Parameter estimation and statistical inference

Suppose that we have an observed time series r_t, $t = -s + 1, ..., T$, where s denotes the required number of initial values. Conditional on the initial values of r_t and h_t the log-likelihood function of the relevant (skewed) GARCH model can be written

---

[16]A detailed discussion on using this distribution in GARCH models can be found in Miettinen (2007).

as

$$l_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} \log f_{t-1}(r_t; \boldsymbol{\theta}),$$

where $f_{t-1}(r_t; \boldsymbol{\theta})$ is the conditional density function of r_t given Ω_{t-1} and $\boldsymbol{\theta}$ is the respective vector of unknown parameters. For instance, in the cases of the NIG or NRG specification $f_{t-1}(r_t; \boldsymbol{\theta})$ is obtained from (5) and (15) or (6) and (17), respectively. In contrast to the sequential approach employed by Aas and Haff (2006, 298), our procedure is closer to that of Mencia and Sentana (2004) as we choose to estimate all the parameters of the model simultaneously by maximizing $l_T(\boldsymbol{\theta})$ using standard numerical algorithms.

The ML estimator of the parameter $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}}$, can be treated as approximately normally distributed with mean value $\boldsymbol{\theta}$ and covariance matrix $-(E\partial^2 l_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}')^{-1}$. Approximate standard errors of the components of $\hat{\boldsymbol{\theta}}$ can be obtained by taking the square roots of the diagonal elements of $-(\partial^2 l_T(\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}')^{-1}$. Likelihood ratio, Wald, and Lagrange multiplier tests with approximate chi-square distributions can also be performed in the usual way.

2.3 Empirical examples

In this subchapter, the relevance of modeling skewness is illustrated by using return series of two stock market indices. The indices are the Standard & Poor's 500 index from 6.1.1986 to 12.31.1997 and the Amsterdam EOE index from the same period. For both data, we use about 3000 first observations for the estimation of the models and save 1000 observations for out-of-sample analysis.[17]

---

[17]Most of the empirical work, including the estimation of the models, was carried out on GAUSS. In the calculation of the modified Bessel function with integer order, formulae 9.8.5 - 9.8.8 from Abramowitz and Stegun (1970) were applied. Bessel functions with fractional order were calculated by using algorithms extracted from Spanier and Oldham (1987; 494-495, 504-505). Numerical integration of the density functions containing a modified Bessel function was executed with Matlab.

2.3.1   S&P 500 index

The return series of the Standard & Poor's 500 index has the stylized facts typical of a financial time series. In particular volatility clustering can be observed from the plot of the series, while the unconditional excess kurtosis is observable in the figure presenting the estimated density and the summary statistics in Table 1.

Table 1: Summary statistics of the returns of the S&P 500 index (estimation period).

Observations	3029
Mean	0.050
Std.devn.	1.009
Skewness	-4.237
Excess kurtosis	93.629
Minimum	-22.833
Maximum	8.709
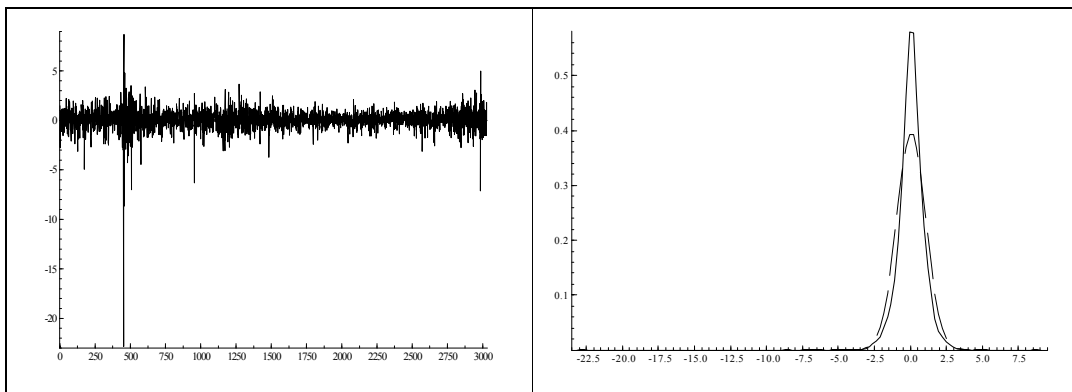
Figure 2. Plot of the returns of the S&P 500 index and estimated density against the density of the normal distribution (dashed lines, estimation period).

We consider the GARCH(1,1)-M model with conditional t distribution as a benchmark for the models with skewed conditional distribution. The estimation results for this model are presented in the first column of Table 2 (degrees of freedom parameter of the t distribution is denoted by df). The specification of the conditional mean is of the form vh_t. Both constant and lagged returns were found to be statistically insignificant in the mean equation. The Ljung-Box statistic for the stan-

dardized residuals also indicated no remaining autocorrelation. The specification of the conditional variance is the standard GARCH(1,1).

Table 2: Estimation results for GARCH(1,1)-M models for the returns of the S&P 500 index.

	t	ST-t	ST-NIG	ST-NRG
ν	0.075	0.099	0.069	0.067
	(0.017)	(0.018)	(0.020)	(0.021)
a_0	0.010	0.014	0.019	0.019
	(0.003)	(0.007)	(0.009)	(0.009)
a_1	0.041	0.033	0.046	0.041
	(0.008)	(0.007)	(0.010)	(0.009)
b_1	0.947	0.856	0.811	0.825
	(0.009)	(0.049)	(0.056)	(0.055)
d_1		0.260	0.300	0.297
		(0.118)	(0.136)	(0.131)
g_1		1.628	1.606	1.635
		(0.116)	(0.147)	(0.131)
κ			0.490	0.429
			(0.140)	(0.144)
α			1.194	
			(0.164)	
β			-0.131	-0.248
			(0.049)	(0.091)
λ				-2.571
				(0.251)
df	4.692	4.773		
	(0.412)	(0.424)		
$\hat{l}$	-3581.6	-3566.0	-3561.9	-3556.3
AIC	7173.2	7146.0	7141.7	7130.7
BIC	7203.3	7188.1	7195.9	7184.7
Skewness	0	0	-0.302	-
Excess kurtosis	8.671	7.762	2.649	-
$Q(20)$	28.920	28.919	27.213	27.640
$Q^2(20)$	20.246	20.978	11.354	13.906

Skewness and excess kurtosis are the implied point estimates of the skewness and excess kurtosis of the standardized error η_t. Q(20) stands for Ljung-Box test for up to 20th-order serial correlations in standardized residuals. $Q^2(20)$ stands for Ljung-Box test for up to 20th-order serial correlations in squared standardized residuals. Standard errors (in parentheses) obtained from inversed Hessian.

From the results we observe that $\hat{a}_1 + \hat{b}_1 = 0.988$, which implies relatively slow decline of the conditional variance.

The second column of Table 2 presents estimation results for the respective ST-GARCH model.[18] We observe that the sum $\hat{a}_1 + \hat{b}_1$ now equals 0.889, which may be considered substantially lower than in the standard GARCH case. It may be noted that the parameters related to the nonlinear part of the model for the conditional variance are estimated precisely, as the standard errors are comparatively low. A substantial improvement in the log-likelihood and thus in the values of the AIC and BIC criteria is also observed. In light of the estimation results it appears that the ST-GARCH specification is preferable to the standard GARCH.

As already mentioned, it has been suggested in the literature that the GH distribution itself may be overparameterized for GARCH applications. Our experience corroborates. When a GARCH-GH model was estimated for the S&P 500 series an insignificant estimate for the parameter α was obtained. In addition, the difference $\hat{\alpha} - \left|\hat{\beta}\right|$ was as small as 0.0001, suggesting that the restriction $\alpha \rightarrow |\beta|$ could be binding. The point estimate of λ was about -2.5, indicating that the restriction $\lambda < 0$ could also be binding. Based on these results NIG and NRG models were fitted. It should be noted that we also tried models with $\lambda = 1$ (hyperbolic distribution) and $\delta \rightarrow 0$ (variance-gamma distribution), but their empirical performance was consistently surpassed by the $\lambda = -1/2$ (NIG) and $\alpha \rightarrow |\beta|$ (NRG) cases. In addition, we considered the z distribution introduced into GARCH modeling by Lanne and Saikkonen (2007). In our empirical applications this distribution behaved very similarly to the NIG distribution.[19]

In the following columns, estimation results for ST-GARCH models with skewed distributions are reported. The results show that both specifications imply negative skewness. For the NIG and NRG distributions the skewness can be interpreted as statistically significant by observing that the estimates of parameter β are statistically significant.

---

[18]The GJR-GARCH specification was not found necessary with the ST-GARCH specification.

[19]For estimation results for variance-gamma and z distribution models see Miettinen (2007).

Table 2 reveals that estimates of the parameters in the conditional mean and conditional variance are not affected by the choice of skewed distributions. Differences in estimates appear to be relatively small in models based on the NIG and NRG distributions. However, compared to the results obtained with the t distribution the sum $\hat{a}_1 + \hat{b}_1$ is further reduced, now being less than 0.85 in the smaller case. Also, estimates of the parameter κ are significantly different from zero in both skewed specifications.

Using the moments presented in subchapter 2.2 with the restrictions from subchapter 2.3, point estimates for the skewness and excess kurtosis of the distribution of the standardized error can be computed. These estimates are also presented in Table 2. In the case of the NRG distribution skewness and (excess) kurtosis do not exist because the estimate of the parameter λ is greater than -3. Of course, the estimated distribution is still negatively skewed as the estimated skewness parameter is negative. In the case of the NIG distribution these quantities do indeed exist, the respective point estimates being -0.302 and 2.649.

According to AIC, the models based on the NIG and NRG distributions are preferable to the models based on the t distribution, and of all the models the one based on the NRG distribution is recommended by both AIC and BIC. However, according to BIC, the t distribution is preferred to the NIG distribution.

From the estimation results we now proceed to a more informal examination of the models. A standard procedure in the diagnostic checking of a GARCH model is to view the distribution of the standardized residuals, which, for a correctly specified model, should be compatible with the distribution assumed for the (theoretical) error term. We examine the distribution of the standardized residuals by comparing the estimated density function with its theoretical counterpart. To make the fit in the tails more easily visible, plots with logarithmic scale are presented. The density of the residuals can be estimated using the standard methods presented, for example, in Silverman (1986). We set the window width of the smoothing parameter equal to $1.06\hat{\sigma}/T^{0.2}$, where $\hat{\sigma}$ denotes the estimated standard deviation of the data, and use the density of the normal distribution as a kernel. In the upper left corner of

Figure 3 the estimated (logarithmic) density of the standardized residuals obtained from the GARCH(1,1)-M-model based on the t distribution is presented against the theoretical distribution implied by the model. The figure reveals that the left tail is constantly underestimated as the estimated density exceeds the theoretical density. The right tail appears to be overestimated for values of densities exceeding 3. The corresponding figure for the ST-GARCH(1,1)-M-t model shows that the addition of the smooth transition term to the conditional variance has not changed the properties of the empirical distribution of the standardized residuals.
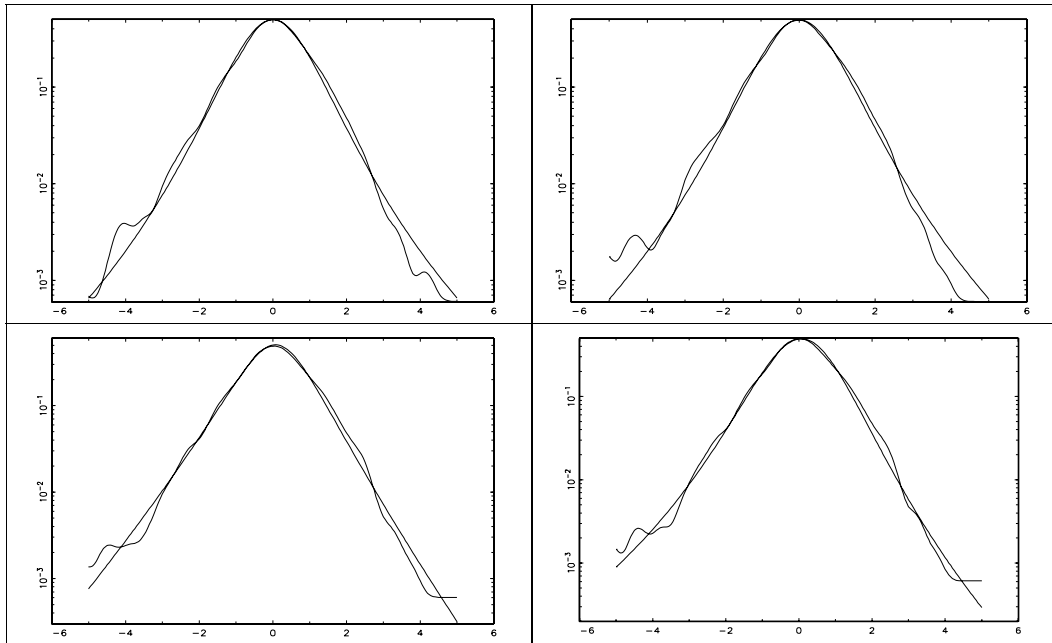
Figure 3. Plots of the logarithmic estimated densities of the residuals of the GARCH(1,1)-M-t (upper left), ST-GARCH(1,1)-M-t (upper right), ST-GARCH(1,1)-M-NIG (lower left) and ST-GARCH(1,1)-M-NRG (lower right) models against the logarithmic theoretical densities.

As to the skewed distributions, Figure 3 clearly shows the conditional skewness, as the left tails of the theoretical densities are thicker than the right tails. The fit in the tails is also clearly better than in the t distribution. The performance of the NRG distribution appears nearly flawless.

From Figure 3 it can be concluded that, using either of the skewed distributions, the fit can be improved especially in the tails of the distribution. In order to illustrate

the potential importance of such an improvement, a Value-at-Risk application[20] is presented.

In Value-at-Risk (VaR) applications the interest is in determining a return $VaR_{t|t-1}$, the smaller of which is observed at some (small) probability, say p. Thus, if the VaR of size p is properly determined, it should be observed that $p \times T$ of T returns are smaller than the VaR of size p. The first line in the left panel of Table 3 displays the percentages of observations in the estimation period that are smaller than the 1, 2, and 3 % VaR for the GARCH model based on the conditional t distribution. In all cases the percentages observed are higher than the size of the VaR. The interpretation is that the t distribution consistently underestimates the left tail of the distribution. The situation is even worse in the case of the ST-GARCH-t model, as indicated by the in-sample figures on the second row of the table.

Table 3: Observed percentages of violations of VaR for different models.

	In-sample VaR			Out-of-sample VaR		
	1 %	2 %	3 %	1 %	2 %	3 %
t	1.22	2.41^{21}	3.23^{22}	1.20	3.40^{23}	4.80^{24}
ST-t	1.25	2.81^{25}	3.37^{26}	1.40	3.30^{27}	5.00^{28}
ST-NIG	0.89	1.78	2.91	1.00	2.10	3.90
ST-NRG	0.86	1.95	3.10^{29}	0.90	2.50	4.50^{30}

The respective figures for the skewed distributions (lines 3-4 in the left panel of Table 3) show that the VaR is not overestimated, but underestimated, by the models based on skewed distributions. Only in one case out of 6 does the observed

---

[20]Various GARCH models have been considered for VaR estimation e.g. by Angelidis et al. (2004).

[21]Rejected at 1 % significance level by Christoffersen's test.

[22]Rejected at 1 % significance level by Christoffersen's test.

[23]Rejected at 1 % significance level by Kupiec's test and at 5 % level by Christoffersen's test.

[24]Rejected at 1 % significance level by Kupiec's and Christoffersen's tests.

[25]Rejected at 5 % significance level by Kupiec's test and at 0.1 % level by Christoffersen's test.

[26]Rejected at 1 % significance level by Christoffersen's test.

[27]Rejected at 1 % significance level by Kupiec's test and at 5 % level by Christoffersen's test.

[28]Rejected at 1 % significance level by Kupiec's and Christoffersen's tests.

[29]Rejected at 5 % significance level by Christoffersen's test.

[30]Rejected at 1 % significance level by Kupiec's test and at 5 % level by Christoffersen's test.

percentage of violations exceed the VaR size. Generally, not only is this underestimation smaller than the overestimation (in terms of percentages), but it is also less likely to be a nuisance in the actual applications as it leads to more conservative VaR figures.

The models considered are tested even harder in out-of-sample VaR analysis using the 1000 last observations of our data. The results are reported in the left panel of Table 3. Models based on t distribution appear to underestimate the left tail of the distribution, also in out-of-sample analysis. Now this is also true for the models based on skewed distributions for VaRs of sizes 2 and 3 %. However, this underestimation is considerably smaller for the skewed models than for the symmetric models.

In order to draw conclusions on the differences in Table 3, a test proposed by Kupiec (1995) is performed. The no rejection regions at the 5 % confidence level for $T = 3029$ are $0.7 \leq 100N/T \leq 1.35$, $1.55 \leq 100N/T \leq 2.51$ and $2.44 \leq 100N/T \leq 3.60$ (where N is the number of violations), for the values of $100p$ equal to 1, 2 and 3 respectively. In the case of the skewed distributions the number of violations obtained from Table 3 falls within these regions so that no VaR can be rejected. However, in out-of-sample VaR forecasting using 1000 last observations the models based on the t distribution are rejected at the 1 % confidence level for $100p = 2$ and $100p = 3$ (right panel of Table 3). The differences in the forecast performance of the models based on the skewed distributions appear to remain relatively small except for the NRG distribution, whose performance deteriorates as the size of VaR increases.

Kupiec's (1995) test only allows us to draw conclusions on the unconditional coverage of the VaR forecasts. A likelihood ratio test designed to account for unconditional coverage and the independence of the failures is presented in Christoffersen (1998). Using this more sophisticated test the 2 % and 3 % VaRs based on the t distribution are rejected at 5 % significance level both in-sample and out-of-sample. However, the VaRs based on the skewed distributions are only rejected in the case of the NRG distribution with $100p = 3$ both in-sample and out-of-sample.

In light of these analyses not much can be said about which skewed distribution one should apply. The plots of the estimated densities suggest that when the goal is to model the conditional distribution as a whole the NRG distribution is the best choice. However, when modeling of the left tail is considered, the situation is less clear, although it seems reasonable to argue that in VaR applications both skewed distributions outperform the t distribution. Obviously the distribution should be selected according to the size of the VaR, as the performance of the distribution appears to have some relation to it.

2.3.2 Amsterdam EOE index

Some of the conclusions drawn on the basis of with the returns of the S&P 500 index can be repeated by analyzing the returns of the Amsterdam EOE index (Figure 4). Summary statistics of the series are presented in Table 4.

Table 4: Summary statistics of the returns of the Amsterdam EOE index (estimation period).

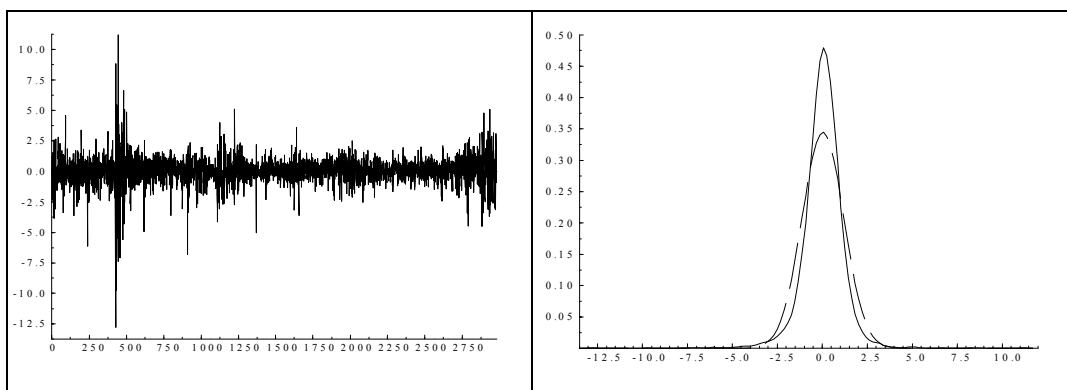|	
Observations	2979
Mean	0.040
Std.devn.	1.158
Skewness	-0.683
Excess kurtosis	15.92
Minimum	-12.788
Maximum	11.179

Figure 4. Plot of the returns of the Amsterdam EOE index and estimated density against the density of the normal distribution (dashed lines, estimation period).

Table 5: Estimation results for GARCH(1,1) models for the returns of the Amsterdam EOE index.

	GJR-t	NIG	NRG
ϕ_0	0.077	0.050	0.051
	(0.015)	(0.016)	(0.016)
a_0	0.021	0.022	0.020
	(0.006)	(0.006)	(0.006)
a_1	0.041	0.071	0.071
	(0.013)	(0.012)	(0.012)
b_1	0.911	0.900	0.904
	(0.016)	(0.017)	(0.016)
c_1	0.051		
	(0.018)		
κ		0.352	0.315
		(0.102)	(0.100)
α		1.924	
		(0.291)	
β		-0.339	-0.552
		(0.100)	(0.171)
λ			-3.379
			(0.378)
df	5.935		
	(0.582)		
$\hat{l}$	-3955.7	-3953.4	-3943.4
AIC	7923.4	7920.8	7900.8
BIC	7959.5	7962.9	7942.9
Skewness	0	-0.384	-0.630
Excess Kurtosis	3.101	1.781	-
$Q(20)$	23.518	23.420	23.516
$Q^2(20)$	7.419	7.389	7.541

Skewness and excess kurtosis are the implied point estimates of the skewness and excess kurtosis of the standardized error η_t. $Q(20)$ stands for Ljung-Box test for up to 20th-order serial correlations in standardized residuals. $Q^2(20)$ stands for Ljung-Box test for up to 20th-order serial correlations in squared standardized residuals. Standard errors (in parentheses) obtained from inversed Hessian.

In this case, the GJR-GARCH specification was found satisfactory in the symmetric case (Table 5). The conditional mean was only modeled with a constant as no GARCH-M effect was discovered.

When fitting a GARCH model with GH innovations the estimation results were qualitatively the same as in the previous example. More precisely, this time we obtained an insignificant estimate of β and the difference $\widehat{\alpha} - \left|\widehat{\beta}\right|$ was 0.0033, which again suggests that the restriction $\alpha \to |\beta|$ could be binding. The point estimate of λ was about -3.0. In light of these results NIG and NRG models were again employed.[31]
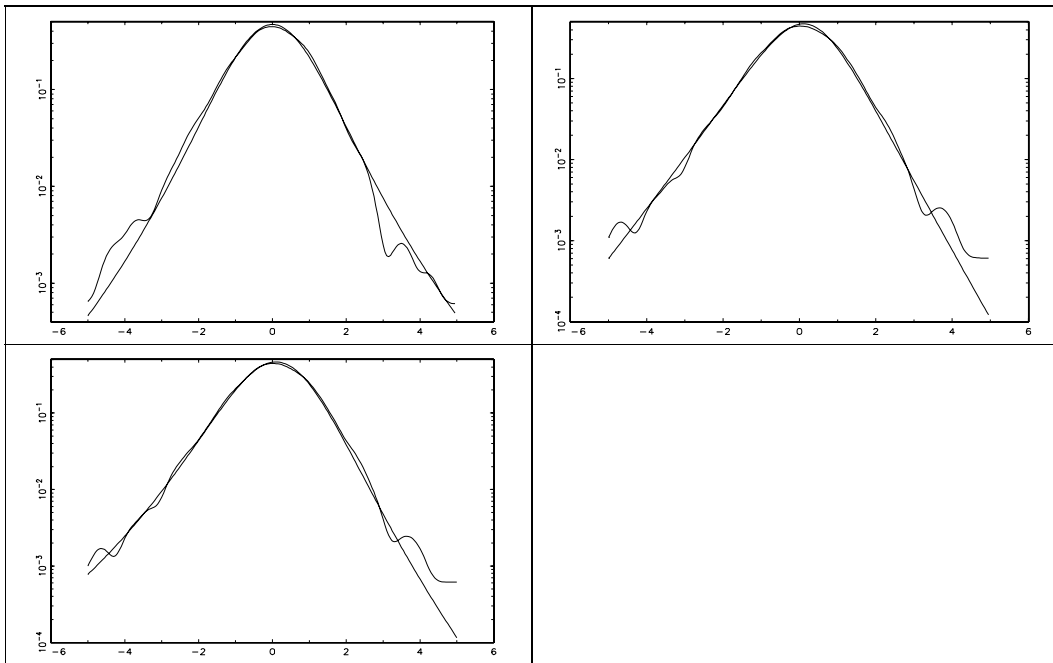

Figure 5. Plots of the logarithmic estimated densities of the residuals of the GJR-GARCH(1,1)-t (upper left), GARCH(1,1)-NIG (upper right) and GARCH(1,1)-NRG (lower left) models against the logarithmic theoretical densities.

The estimation of models employing skewed conditional distributions was carried out successfully. The GJR specification was not found necessary when the extension from Equation (12) was added to the conditional variance equation. The estimates of the parameters of the NIG distribution again imply significant negative skewness of -0.384. In the case of the NRG distribution the point estimate of the skewness is -0.63. A possible explanation for such a large estimate is that the estimate of λ

---

[31]We also tried the $\lambda = 1$ and $\delta \to 0$ cases practically repeating the results from the previous example. The same applies to the z distribution.

is relatively close to -3, the point at which the skewness no longer exists. Again, the model based on the NRG distribution is recommended by AIC and BIC.

Figures of estimated densities of the error term and the theoretical densities (Figure 5) suggest that the fit is considerably improved in the tails by allowing for conditional skewness.

Finally a VaR example is constructed in order to illustrate the potential advantages of allowing for conditional skewness. The results are presented in Table 6. The first line of the left panel suggests that the application of the t distribution leads to too many violations. Skewed distributions appear to lead to too small VaRs which, as already mentioned, may be in practice a less harmful feature. The out-of-sample results (right panel) also indicate very good performance by the models based on the skewed distributions and considerable underestimation of the left tail by the model based on the t distribution. Unfortunately the formal tests applied in the previous subchapter do not allow us to reject any models in this case. It may be noted that, again, it is difficult to say which skewed distribution one should apply.

Table 6: Observed percentages of violations of VaR for different conditional distributions.

	In-sample VaR			Out-of-sample VaR		
	1 %	2 %	3 %	1 %	2 %	3 %
t	1.28	2.42	3.46	1.60	2.70	4.10
NIG	0.73	1.81	2.72	1.00	1.70	3.00
NRG	0.77	1.88	2.89	1.10	1.80	3.30

2.4 Discussion

This chapter discussed the importance of accommodating conditional skewness in GARCH modeling. This issue was addressed by applying skewed distributions as a conditional distribution of the GARCH model. The emphasis was in the empirical performance of the models considered.

Specifically, we considered so-called generalized hyperbolic distributions for GARCH models. Some of the special cases of these distributions have been used

previously, but this thesis may be one of the first attempts to use the GH distribution. However, the estimation was found to converge to the special case, where the mixing variable in the normal variance-mean mixture representation of the GH distribution has a reciprocal or inverse gamma distribution.

In our empirical applications to returns of two real stock market indices, statistically significant estimates of the parameters of the skewed distributions employed were obtained. In particular, all the estimates implied negatively skewed conditional distributions with point estimates of the conditional skewness, when attainable, falling between approximately -0.3 and -0.6. Compared to models based on the conditional t distribution, the skewed models were also supported by model selection criteria. Furthermore, the estimated densities of standardized residuals implied by the models revealed that the t distribution may not be adequate if the objective is to model the tails carefully. In this respect, the skewed distributions employed performed considerably better.

To demonstrate the potential importance of the improved fit in the tails of the distribution, VaR applications were performed. The t distribution consistently evaluated the left tail of the distribution too thin so that the probability of small returns was underestimated. By contrast, with skewed distributions probabilities of small returns seemed to be slightly overestimated rather than underestimated. This means specifically that when the objective is to avoid modeling the left tail too liberally it is preferable to use skewed distributions instead of the t distribution.

As discussed in the introduction, future work on GARCH-GH models includes options pricing when the underlying asset follows a GARCH-GH model. The marginal effect of the generalized hyperbolic specification against alternatives such as mixture GARCH models should also be more carefully addressed.

References

Aas, K. & I. H. Haff 2006: The generalized hyperbolic skew Student's t-distribution, *Journal of Financial Econometrics*, 4, 275-309.

Abramovitz, M. & I. A. Stegun 1970: Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Washington, D.C.: Knovel.

Angelidis, T., A. Benos & S. Degiannakis 2004: The use of GARCH models in VaR estimation, *Statistical Methodology*, 1, 105-128.

Barndorff-Nielsen, O. E. 1997: Normal inverse Gaussian distributions and stochastic volatility modeling, *Scandinavian Journal of Statistics*, 24, 1-13.

Barndorff-Nielsen, O. E., J. Kent & M. Sørensen 1982: Normal variance-mean mixtures and z distributions, *International Statistical Review*, 50, 145-159.

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

Carrasco, M. & X. Chen 2002: Mixing and moment properties of various GARCH and stochastic volatility models, *Econometric Theory*, 18, 17-39.

Christoffersen, P. F. 1998: Evaluating interval forecasts, *International Economic Review*, 39, 841-862.

Eberlein, E. 2001: Application of generalized hyperbolic Lévy motions to finance, in O. E. Barndorff-Nielsen, T. Mikosch & S. I. Resnick, ed., Lévy Processes: Theory and Applications, Boston: Birkhäuser Verlag, 319-337.

Eberlein, E. & E. A. v. Hammerstein 2003: Generalized hyperbolic and inverse Gaussian distributions: Limiting cases and approximation of processes, Proceedings of the 4th Ascona Conference, Birkhäuser Verlag.

Glosten, L. R., R. Jagannathan & D. E. Runkle 1993: On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance*, 48, 1779-1801.

Jensen, M. B. & A. Lunde 2001: The NIG-S&ARCH model: a fat-tailed, stochastic, and autoregressive conditional heteroskedastic volatility model, *Econometrics Journal*, 4, 319-342.

Jørgensen, B. 1982: Statistical Properties of the Generalized Inverse Gaussian Distribution. New York: Springer-Verlag.

Kupiec, P. H. 1995: Techniques for verifying the accuracy of risk measurement models, *The Journal of Derivatives*, 3, 73-84.

Lanne, M. & P. Saikkonen 2005: Non-linear GARCH models for highly persistent volatility, *Econometrics Journal*, 8, 251-276.

Lanne, M. & P. Saikkonen 2007: Modeling conditional skewness in stock returns, *European Journal of Finance*, 13, 691-704.

Madan, D. B., P. P. Carr & E. C. Chang 1998: The variance gamma process and option pricing, *European Finance Review*, 2, 79-105.

Meitz, M. & P. Saikkonen 2008: Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD Models, *Econometric Theory*, 24, 1291-1320.

Mencia, F. J. & E. Sentana 2004: Estimation and testing of dynamic models with generalized hyperbolic innovations, CEMFI Working Paper, 0411.

Miettinen, J. 2007: GARCH modeling with normal variance-mean mixtures. Licentiate Thesis, University of Helsinki.

Prause, K. 1999: The generalized hyperbolic model: Estimation, financial derivatives, and risk measures. Dissertation zur Erlangung des Doktogrades der Mathematischen Fakultät der Albert-Ludwigs-Universität Freiburg i. Br.

Silverman, B. W. 1986: Density Estimation for Statistics and Data Analysis. London: Chapman and Hall.

Spanier, J. & K. B. Oldham 1987: Atlas of Functions. Washington, D.C.: Hemisphere.

3 Mixture MEM models and conditional autoregressive range

This chapter considers mixture multiplicative error models (MMEM). In subchapter 3.1 we present a condition for the strict stationarity of an autoregressive model with non-negative *i.i.d.* coefficients. The discussion is based on Bougerol and Picard (1992a). The strict stationarity condition of the MMEM model is given in subchapter 3.2. The result is obtained by representing the MMEM model as a first order autoregressive process with random coefficients. Because the strict stationarity does not guarantee the existence of second moments, the condition for the weak stationarity of the MMEM model is derived in subchapter 3.3[32]. This condition is applied in subchapter 3.4, where an expression for the autocorrelation function of the MMEM model is derived. In subchapter 3.5 we present an application of the proposed model to the modeling of the conditional autoregressive range. Subchapter 3.6 concludes with discussion.

3.1 Autoregressive model with non-negative *i.i.d.* coefficients

Define $\mathbb{M}^+(d)$ [resp. $(\mathbb{R}^+)^d$] as the set of $d \times d$ matrices (resp. d-dimensional vectors) with non-negative elements, and consider the generalized autoregressive equation

$$X_n = A_n X_{n-1} + B_n, \quad n \in \mathbb{Z}. \tag{18}$$

Here $\left\{ \begin{bmatrix} A_n & B_n \end{bmatrix}', \ n \in \mathbb{Z} \right\}$ is a given sequence of independent, identically distributed (*i.i.d.*), random variables with values in $\mathbb{M}^+(d) \times (\mathbb{R}^+)^d$ and X_n is in $\mathbb{R}^d$. In what follows a necessary and sufficient condition for the existence of a strictly stationary, non-negative, solution of (18) is discussed.

Let us first recall the definition of strict stationarity. A process $\{X_n\}$ is strictly stationary if for all $n, m \in \mathbb{Z}$, the law of $\begin{bmatrix} X_n & X_{n+1} & \dots & X_{n+m} \end{bmatrix}'$ is independent

---

[32]In the preceding literature, the weak stationarity of the mixture GARCH model has been studied by Haas et al. (2004) whereas Liu (2006) considers the weak and the strict stationarity of the Markov-switching GARCH model.

of n. The stationarity properties of the process to be considered later in the chapter are directly related to the stationary solutions of (18). In order to present the stationarity result, the so-called top Lyapunov exponent needs to be defined.

The top Lyapunov exponent is defined in Bougerol and Picard (1992a, 116-117) as follows. Let $\|\cdot\|$ denote any norm on $\mathbb{R}^d$. Then define an operator norm on the set $\mathbb{M}(d)$ of $d \times d$ matrices by

$$\|M\| = \sup \left\{ \|Mx\| / \|x\| \,;\; x \in \mathbb{R}^d,\ x \neq 0 \right\},$$

for any M in $\mathbb{M}(d)$. The top Lyapunov exponent associated with a sequence $\{A_n, n \in \mathbb{Z}\}$ of $i.i.d.$ random matrices, is defined by

$$\gamma = \inf \left\{ E \left(\frac{1}{n+1} \log \|A_0 A_{-1} \ldots A_{-n}\| \right),\ n \in \mathbb{N} \right\},$$

when $E(\log^+ \|A_0\|)$ is finite $[\log^+ x = \max(\log x, 0)]$. We point out that, almost surely,

$$\gamma = \lim_{n \to +\infty} \frac{1}{n} \log \|A_0 A_{-1} \ldots A_{-n}\|.$$

This result allows us to estimate γ by simulation. It may also be noted that γ is independent of the norm.

Now suppose that in (18) $E(\log^+ \|A_n\|)$ is finite and that all the coefficients of A_n (or B_n) are strictly positive with nonzero probability. Then, according to Theorem 3.2 in Bougerol and Picard (1992a, 123), if Equation (18) has a strictly stationary, non-negative, solution, then the top Lyapunov exponent γ related to the matrices $\{A_n,\ n \in \mathbb{Z}\}$ is strictly negative. Moreover, if $E(\log^+ \|B_0\|)$ is finite and if $\gamma < 0$, then for all $n \in \mathbb{Z}$, the series

$$X_n = B_n + \sum_{k=1}^{\infty} A_n A_{n-1} \ldots A_{n-k+1} B_{n-k}$$

converges almost surely and the process $\{X_n,\ n \in \mathbb{Z}\}$ is the unique strictly stationary and ergodic solution of (18). In Bougerol and Picard (1992a) the result pre-

sented above is applied to develop conditions for the strict stationarity of GARCH processes. This application is presented in Appendix 1.

It may be noted that if the assumption of the non-negativeness of $\left\{ \left[\begin{array}{cc} A_n & B_n \end{array} \right]', \ n \in \mathbb{Z} \right\}$, or, especially in our applications, the non-negativeness of $\{A_n, \ n \in \mathbb{Z}\}$, is relaxed, the condition for the strict stationarity of $\{X_n, \ n \in \mathbb{Z}\}$ can be concluded from Theorem 2.5 of Bougerol and Picard (1992b, 1717-1718). In particular, the non-negativeness of $\{A_n, \ n \in \mathbb{Z}\}$ can be relaxed if we wish to allow the parameters of the MEM model (see following subchapter) to satisfy the condition given in Nelson and Cao (1992).

In subchapter 3.3 a condition for the weak stationarity for the MMEM model is obtained by presenting the MMEM model as a special case of the generalized AR model (18). This result is applied in subchapter 3.4, where an expression for the autocorrelation function for this model is derived.

3.2   Strict stationarity of the mixture MEM model

In Engle (2002) the multiplicative error model (MEM) is specified as

$$x_t = \mu_t \varepsilon_t,$$

where $\varepsilon_t \sim D(1, \phi^2)$, that is, ε_t has a non-negative distribution with unit mean and variance ϕ^2. The innovations ε_t are also assumed to be independently and identically distributed. The equation for the conditional expectation of x_t given the information set $\Omega_{t-1} = \{x_{t-1}, x_{t-2}, ...\}$ is generally specified as

$$\mu_t = \omega + \sum_{i=1}^{q} \alpha_i x_{t-i} + \sum_{j=1}^{p} \beta_j \mu_{t-j}.$$

In order to guarantee the non-negativity of the mean process $\{\mu_t\}$, the parameters $\alpha_i, \ i = 1, ..., q$ and $\beta_j, \ j = 1, ..., p$, are supposed to satisfy the conditions given in Nelson and Cao (1992). The MEM models nests the squared (that is, $y_t^2 = h_t v_t^2$, where $v_t^2 | \Omega_{t-1} \sim D(1, \phi_t^2)$) GARCH model by Bollerslev (1986), the ACD model by Engle and Russel (1998) as well as the CARR model by Chou (2005).

Now consider the generalization of the MEM model given by

$$x_t = \sum_{i=1}^{m} I(\eta_t = i)\mu_{i,t}\varepsilon_{i,t} \tag{19}$$

where

$$\mu_{i,t} = \omega_i + \sum_{j=1}^{q} \alpha_{i,j}x_{t-j} + \sum_{l=1}^{m}\sum_{k=1}^{p} \beta_{i,l,k}\mu_{l,t-k}, \tag{20}$$

and η_t and $\varepsilon_{i,t}$ are independently and identically distributed with $P(\eta_t = i) = \pi_i$, $i = 1, ..., m$, $\sum_{i=1}^{m} \pi_i = 1$, and $\varepsilon_{i,t} \sim D_i(1, \phi_i^2)$. Moreover, the processes $\{\varepsilon_{i,t}\}$ and $\{\eta_t\}$ are mutually independent for all i. In order to be able to apply the results given in the previous subchapter, we assume that all the coefficients ω_i, $\alpha_{i,j}$ and $\beta_{i,l,k}$ are non-negative. If the model is restricted to the diagonal model, that is $l = i$ for $i = 1, ..., m$, the conditions given in Nelson and Cao (1992) can be employed.

This model, as an obvious extension of the MEM model, nests some previously considered time series models. As special cases the (finite) mixture ACD model by De Luca and Zuccolotto (2003), the MN-GARCH model by Haas et al. (2004) and Alexander and Lazar (2006), and the mixture MEM model by Lanne (2006) can be mentioned. Our aim is to present the model in convenient form, so that the results of Bougerol and Picard (1992a) can be directly applied to obtain a condition for the strict stationarity for these models.

Now, let us define the row vectors

$$
\begin{aligned}
\boldsymbol{\mu}_{i,t} &= \left[\begin{array}{ccc} \mu_{i,t} & \cdots & \mu_{i,t-p+1} \end{array} \right] \in \mathbb{R}^p, \\
\tau_{i,t} &= \left[\begin{array}{cccc} \beta_{i,i,1} + \alpha_{i,1}I(\eta_{t-1} = i)\varepsilon_{i,t-1} & \beta_{i,i,2} & \cdots & \beta_{i,i,p-1} \end{array} \right] \in \mathbb{R}^{p-1}, \\
\xi_{i,j,t} &= \left[\begin{array}{cccc} \beta_{i,j,1} + \alpha_{i,1}I(\eta_{t-1} = j)\varepsilon_{j,t-1} & \beta_{i,j,2} & \cdots & \beta_{i,j,p-1} \end{array} \right] \in \mathbb{R}^{p-1}, \ i \neq j, \\
\xi_{i,t} &= \left[\begin{array}{cccc} I(\eta_{t-1} = i)\varepsilon_{i,t-1} & 0 & \cdots & 0 \end{array} \right] \in \mathbb{R}^{p-1}, \\
\alpha_i &= \left[\begin{array}{ccc} \alpha_{i,2} & \cdots & \alpha_{i,q-1} \end{array} \right] \in \mathbb{R}^{q-2}, \\
\delta_i &= \left[\begin{array}{cccc} \omega_i & 0 & \cdots & 0 \end{array} \right] \in \mathbb{R}^p.
\end{aligned}
$$

The definitions

$$X_t = \begin{bmatrix} \boldsymbol{\mu}_{1,t} & \boldsymbol{\mu}_{2,t} & \cdots & \boldsymbol{\mu}_{m,t} & x_{t-1} & \cdots & x_{t-q+1} \end{bmatrix}' \in \mathbb{R}^{mp+q-1}$$

and

$$B = \begin{bmatrix} \delta_1 & \delta_2 & \cdots & \delta_m & 0 & \cdots & 0 \end{bmatrix}' \in \mathbb{R}^{mp+q-1}$$

coupled with the definition of the $(mp + q - 1) \times (mp + q - 1)$ matrix A_t, written in block form, as

$$A_t = \begin{bmatrix}
\tau_{1,t} & \beta_{1,1,p} & \xi_{1,2,t} & \beta_{1,2,p} & \cdots & \xi_{1,m,t} & \beta_{1,m,p} & \alpha_1 & \alpha_{1,q} \\
I_{p-1} & 0 & 0 & 0 & & 0 & 0 & 0 & 0 \\
\xi_{2,1,t} & \beta_{2,1,p} & \tau_{2,t} & \beta_{2,2,p} & & \xi_{2,m,t} & \beta_{2,m,p} & \alpha_2 & \alpha_{2,q} \\
0 & 0 & I_{p-1} & 0 & & 0 & 0 & 0 & 0 \\
\vdots & & & & \ddots & & & & \vdots \\
\xi_{m,1,t} & \beta_{m,1,p} & \xi_{m,2,t} & \beta_{m,2,p} & & \tau_{m,t} & \beta_{m,m,p} & \alpha_m & \alpha_{m,q} \\
0 & 0 & 0 & 0 & & I_{p-1} & 0 & 0 & 0 \\
\xi_{1,t} & 0 & \xi_{2,t} & 0 & & \xi_{m,t} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \cdots & 0 & 0 & I_{q-2} & 0
\end{bmatrix},$$

allow us to state the model defined by Equations (19) and (20) in the convenient form

$$X_t = A_t X_{t-1} + B.$$

Now, according to Theorem 3.2 in Bougerol and Picard (1992a) the mixture MEM model is strictly stationary and ergodic, if the top Lyapunov exponent γ associated with matrices $\{A_t,\ t \in \mathbb{Z}\}$ is strictly negative. For a special case of a two component model, see Appendix 2.

3.3 Weak stationarity of the mixture MEM model

In the preceding subchapter a condition for the strict stationarity of the mixture MEM model was presented. However, strict stationarity does not guarantee the existence of second moments. Second moments are often objects of interest, for instance, when considering the autocorrelation structure of the time series modeled.

The existence of second moments follows from the weak stationarity of the process. In this subchapter, a condition for the weak stationarity of the mixture MEM model is derived. The derivation is based on the representation of the mixture MEM model as a special case of the autoregressive model (18).

Again, consider the model (19) but, for simplicity[33], assume that instead of (20) we have

$$\mu_{i,t} = \omega_i + \sum_{j=1}^{q} \alpha_{i,j} x_{t-j} + \sum_{k=1}^{p} \beta_{i,k} \mu_{i,t-k}. \tag{21}$$

Here all the coefficients ω_i, $\alpha_{i,j}$ and $\beta_{i,k}$ are assumed to fulfill the conditions given in Nelson and Cao (1992) to guarantee the positivity of the mean processes $\mu_{i,t}$ for every i.

Then, define

$$
\begin{aligned}
\boldsymbol{\mu}_{i,t} &= \begin{bmatrix} \mu_{i,t} & \cdots & \mu_{i,t-p+1} \end{bmatrix} \in \mathbb{R}^p, \\
\tau_{i,t} &= \begin{bmatrix} \beta_{i,1} + \alpha_{i,1} I(\eta_{t-1}=i)\varepsilon_{i,t-1} & \beta_{i,2} & \cdots & \beta_{i,p-1} \end{bmatrix} \in \mathbb{R}^{p-1}, \\
\xi_{i,j,t} &= \begin{bmatrix} \alpha_{i,1} I(\eta_{t-1}=j)\varepsilon_{j,t-1} & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{p-1}, \\
\xi_{i,t} &= \begin{bmatrix} I(\eta_{t-1}=i)\varepsilon_{i,t-1} & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{p-1}, \\
\alpha_i &= \begin{bmatrix} \alpha_{i,2} & \cdots & \alpha_{i,q-1} \end{bmatrix} \in \mathbb{R}^{q-2}, \\
X_t &= \begin{bmatrix} \boldsymbol{\mu}_{1,t} & \boldsymbol{\mu}_{2,t} & \cdots & \boldsymbol{\mu}_{m,t} & x_{t-1} & \cdots & x_{t-q+1} \end{bmatrix}' \in \mathbb{R}^{mp+q-1}, \\
\delta_i &= \begin{bmatrix} \omega_i & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^p, \\
B &= \begin{bmatrix} \delta_1 & \delta_2 & \cdots & \delta_m & 0 & \cdots & 0 \end{bmatrix}' \in \mathbb{R}^{mp+q-1}.
\end{aligned}
$$

---

[33]We consider the diagonal model to be most relevant in our applications as earlier literature (see e.g. Haas et al. 2004, 239) suggests only negligible improvement in fit by allowing nondiagonal coefficients $\beta_{i,l,k}, i \neq l$ to differ from zero.

Using these definitions, the $(mp + q - 1) \times (mp + q - 1)$ matrix A_t becomes

$$A_t = \begin{bmatrix} \tau_{1,t} & \beta_{1,p} & \xi_{1,2,t} & 0 & \cdots & \xi_{1,m,t} & 0 & \alpha_1 & \alpha_{1,q} \\ I_{p-1} & 0 & 0 & 0 & & 0 & 0 & 0 & 0 \\ \xi_{2,1,t} & 0 & \tau_{2,t} & \beta_{2,p} & & \xi_{2,m,t} & 0 & \alpha_2 & \alpha_{2,q} \\ 0 & 0 & I_{p-1} & 0 & & 0 & 0 & 0 & 0 \\ \vdots & & & & \ddots & & & & \vdots \\ \xi_{m,1,t} & 0 & \xi_{m,t} & 0 & & \tau_{m,t} & \beta_{m,p} & \alpha_m & \alpha_{m,q} \\ 0 & 0 & 0 & 0 & & I_{p-1} & 0 & 0 & 0 \\ \xi_{1,t} & 0 & \xi_{2,t} & 0 & & \xi_{m,t} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & I_{q-2} & 0 \end{bmatrix},$$

which allows us to write

$$X_t = A_t X_{t-1} + B.$$

Note, that the matrix A_t can be rewritten as

$$A_t = \sum_{i=1}^{m} I(\eta_t = i) \left(\alpha e_{i,t-1} + C \right),$$

where

$$\alpha = \begin{bmatrix} \alpha_{1,1} & 0_{(1 \times (p-1))} & \alpha_{2,1} & 0_{(1 \times (p-1))} & \cdots & \alpha_{m,1} & 0_{(1 \times (p-1))} & 1 & 0_{(1 \times (q-2))} \end{bmatrix}'$$
$$\in \mathbb{R}^{mp+q-1},$$

$$e_{i,t-1} = \begin{bmatrix} 0_{(1 \times (i-1)p)} & \varepsilon_{i,t-1} & 0_{(1 \times ((m-i)p+p+q-2))} \end{bmatrix} \in \mathbb{R}^{mp+q-1},$$

with $0_{(1 \times i)}$ denoting a $1 \times i$ vector of zeros, and

$$
C = \begin{bmatrix}
\beta_1 & \beta_{1,p} & 0 & 0 & \cdots & 0 & 0 & \alpha_1 & \alpha_{1,q} \\
I_{p-1} & 0 & 0 & 0 & & 0 & 0 & 0 & 0 \\
0 & 0 & \beta_2 & \beta_{2,p} & & 0 & 0 & \alpha_2 & \alpha_{2,q} \\
0 & 0 & I_{p-1} & 0 & & 0 & 0 & 0 & 0 \\
\vdots & & & & \ddots & & & & \vdots \\
0 & 0 & 0 & 0 & & \beta_m & \beta_{m,p} & \alpha_m & \alpha_{m,q} \\
0 & 0 & 0 & 0 & & I_{p-1} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \cdots & 0 & 0 & I_{q-2} & 0
\end{bmatrix},
$$

with

$$
\beta_i = \begin{bmatrix} \beta_{i,1} & \cdots & \beta_{i,p-1} \end{bmatrix} \in \mathbb{R}^{p-1}.
$$

The model can now be written as

$$
X_t = \left(\sum_{i=1}^{m} I(\eta_t = i) \left(\alpha e_{i,t-1} + C \right) \right) X_{t-1} + B,
$$

and assuming $\gamma < 0$, after recursive substitution

$$
\begin{aligned}
X_t &= B + \sum_{k=1}^{\infty} A_t A_{t-1} \ldots A_{t-k+1} B \\
&= B + \left(\sum_{i=1}^{m} I(\eta_t = i) \left(\alpha e_{i,t-1} + C \right) \right) B \\
&\quad + \left(\sum_{i=1}^{m} I(\eta_t = i) \left(\alpha e_{i,t-1} + C \right) \right) \left(\sum_{i=1}^{m} I(\eta_t = i) \left(\alpha e_{i,t-1} + C \right) \right) B + \cdots (22)
\end{aligned}
$$

where the series converges almost surely. In what follows, we state a condition for the quadratic convergence of (22), and thus the existence of second moments. The technical details of the proof can be found in Appendix 3.

The mixture MEM model is weakly stationary if the spectral radius of the matrix

$\sum_{i=1}^{m} \pi_i E \left(\left(e'_{i,n-1} \alpha' + C' \right) \otimes \left(e'_{i,n-1} \alpha' + C' \right) \right) = E(A'_n \otimes A'_n)$ is smaller than one. This is written shortly as $\rho(E(A'_n \otimes A'_n)) < 1$.

For a two-component special case, see Appendix 4.

3.4 Autocorrelation structure of the mixture MEM model

Consider the mixture MEM model defined by (19) and (21) and use the notation from the previous subchapter to write it in matrix form as $X_t = A_t X_{t-1} + B$. Assume weak stationarity and let $EX_t = X$ and $EA_t = A$ so that by taking the expectations

$$EX_t = EA_t EX_{t-1} + B \Leftrightarrow X = (I - A)^{-1} B,$$

when $(I - A)^{-1}$ exists. Then by recursive substitution

$$
\begin{aligned}
Cov(X_t, X_{t-k}) &= EX_t X'_{t-k} - XX' \\
&= E(A_t X_{t-1} + B)X'_{t-k} - XX' \\
&= AEX_{t-1}X'_{t-k} + BEX'_{t-k} - XX' \\
&= AE(A_{t-1}X_{t-2} + B)X'_{t-k} + BX' - XX' \\
&= A^2 EX_{t-2}X'_{t-k} + ABX' + BX' - XX' \\
&\ \ \vdots \\
&= A^k EX_{t-k}X'_{t-k} + \sum_{i=1}^{k} A^{k-i} BX' - XX' \\
&= A^k EX_t X'_t + \sum_{i=1}^{k} A^{k-i} BX' - XX',
\end{aligned}
$$

where the last equality follows from the weak stationarity. The covariance structure of the model is obtained from this after solving $EX_t X'_t$:

$$
\begin{aligned}
EX_t X'_t &= E\left(A_t X_{t-1} + B \right)\left(A_t X_{t-1} + B \right)' \\
&= EA_t X_{t-1}X'_{t-1}A'_t + BX'A' + AXB' + BB'.
\end{aligned}
$$

After using the vectorization operator we have

$$
\begin{aligned}
vecEX_tX_t' &= vecEA_tX_{t-1}X_{t-1}'A_t' + vecBX'A' + vecAXB' + vecBB' \\
&= E\left(A_t \otimes A_t\right)vecEX_{t-1}X_{t-1}' + vecBX'A' + vecAXB' + vecBB' \\
&= E\left(A_t \otimes A_t\right)vecEX_tX_t' + vecBX'A' + vecAXB' + vecBB'
\end{aligned}
$$

or

$$
(I_{(mp+q-1)^2} - E\left(A_t \otimes A_t\right))vecEX_tX_t' = vecBX'A' + vecAXB' + vecBB'.
$$

By multiplying from the left with the matrix $(I_{(mp+q-1)^2} - E\left(A_t \otimes A_t\right))^{-1}$, which exists when $\rho(E(A_n' \otimes A_n')) < 1,$ we obtain

$$
vecEX_tX_t' = (I_{(mp+q-1)^2} - E\left(A_t \otimes A_t\right))^{-1}\left(vecBX'A' + vecAXB' + vecBB'\right).
$$

3.5 Application to CARR modeling

We consider an application of the mixture MEM models to modeling the (daily) price ranges of financial assets. Let P_t be the logarithmic price of the asset observed at time t, $t = 1, 2, ..., T$. Let P_t^{HIGH} and P_t^{LOW} be the high and low prices between $t-1$ and t and define the range variable as in Chou (2006) by

$$
R_t = P_t^{HIGH} - P_t^{LOW}. \tag{23}
$$

The conditional autoregressive range (CARR) model is specified as

$$
\begin{aligned}
R_t &= \lambda_t \varepsilon_t, \\
\lambda_t &= \omega + \sum_{i=1}^{q} \alpha_i R_{t-i} + \sum_{j=1}^{p} \beta_i \lambda_{t-j}, \\
\varepsilon_t &\sim Gamma(\gamma, \delta).
\end{aligned}
$$

Here the gamma distribution is assumed for the independently and identically dis-

tributed errors. This choice is supported by the earlier empirical evidence (see, for example, Brunetti and Lildholdt 2007). In order to have an error term with unit expected value the restriction $\gamma = 1/\delta$ is employed. The conditional distribution of R_t given its past values is

$$f_{t-1}(R_t) = \frac{1}{\lambda_t \Gamma(\gamma) \delta^\gamma} \left(\frac{R_t}{\lambda_t} \right)^{\gamma-1} \exp \left(-\frac{R_t}{\delta \lambda_t} \right).$$

We call the resulting model CARR-G.

Let us now extend the model to the mixture CARR (MCARR) model. Following Lanne (2006) we consider the two-component case given by

$$R_t = I(\eta_t = 1)\lambda_{1t}\varepsilon_{1t} + (1 - I(\eta_t = 1))\lambda_{2t}\varepsilon_{2t},$$

where the $i.i.d.$ errors $\varepsilon_{it} \sim Gamma(\gamma_i, \delta_i)$, $\gamma_i > 0$, $\delta_i > 0$, $i = 1, 2$, are independent of η_t. The restriction $\gamma_i = 1/\delta_i$ is set for $i = 1, 2$ guaranteeing that $E(\varepsilon_{1t}) = E(\varepsilon_{2t}) = 1$. We also have $\eta_t = 1$ with probability π. The mean equation is specified for the first component by

$$\lambda_{1t} = \omega_1 + \sum_{i=1}^{q} \alpha_{1i} R_{t-i} + \sum_{j=1}^{p} \beta_{1i} \lambda_{1t-j}$$

and for the second component by

$$\lambda_{2t} = \omega_2 + \sum_{i=1}^{q} \alpha_{2i} R_{t-i} + \sum_{j=1}^{p} \beta_{2i} \lambda_{2t-j}.$$

It is easily seen that the conditional expectation of R_t given its past values is

$$
\begin{aligned}
E_{t-1}(R_t) &= E_{t-1}(I(\eta_t = 1)\lambda_{1t}\varepsilon_{1t} + (1 - I(\eta_t = 1))\lambda_{2t}\varepsilon_{2t}) \\
&= \pi E_{t-1}(\lambda_{1t}\varepsilon_{1t}) + (1 - \pi)E_{t-1}(\lambda_{2t}\varepsilon_{2t}) \\
&= \pi\lambda_{1t} + (1 - \pi)\lambda_{2t}.
\end{aligned}
$$

Following from the specified distribution, the conditional distribution of R_t is

$$
\begin{aligned}
f_{t-1}(R_t;\theta) \;=\; & \pi \frac{1}{\lambda_{1t}\Gamma(\gamma_1)\delta_1^{\gamma_1}} \left(\frac{R_t}{\lambda_{1t}}\right)^{\gamma_1-1} \exp\left(-\frac{R_t}{\delta_1\lambda_{1t}}\right) \\
& +(1-\pi)\frac{1}{\lambda_{2t}\Gamma(\gamma_2)\delta_2^{\gamma_2}} \left(\frac{R_t}{\lambda_{2t}}\right)^{\gamma_2-1} \exp\left(-\frac{R_t}{\delta_2\lambda_{2t}}\right).
\end{aligned}
$$

In contrast to the existing literature, in our applications the assumption of the gamma distributed errors appeared to be inadequate. We therefore propose the so-called inverse gamma distribution[34], which, as will be seen, performs considerably better. Under this assumption the conditional distribution of R_t is

$$
\begin{aligned}
f_{t-1}(R_t;\theta) \;=\; & \pi \frac{(\lambda_{1t}\alpha_1)^{\gamma_1}}{\Gamma(\gamma_1)} R_t^{-\gamma_1-1} \exp\left(-\frac{\alpha_1\lambda_{1t}}{R_t}\right) \\
& +(1-\pi)\frac{(\lambda_{2t}\alpha_2)^{\gamma_2}}{\Gamma(\gamma_2)} R_t^{-\gamma_2-1} \exp\left(-\frac{\alpha_2\lambda_{2t}}{R_t}\right),
\end{aligned}
$$

where $\alpha_i = \gamma_i - 1$, $\gamma_i > 0$, $i = 1, 2$, in order to guarantee unit expected value for the errors ε_{it}, $i = 1, 2$. This model will be called the MCARR-IG model.

The estimation of both the standard and mixture CARR models is carried out by using standard numerical methods to maximize the logarithmic likelihood function

$$
l_T(\theta) = \sum_{t=1}^{T} \ln[f_{t-1}(R_t)].
$$

It is well-known that the standard test theory breaks down when determining the number of mixture components (see e.g. Haas et al. 2004, 224 and references therein). As already mentioned, we follow Lanne (2006) and restrict the number of mixture components to two. Other than this, Ahoniemi and Lanne (2009) pointed out that assuming that r_t (or more generally x_t) is stationary and ergodic, it is

---

[34]We also considered the so-called inverse Gaussian, reciprocal inverse Gaussian and hyperbola distributions. None of these distributions performed better than the inverse gamma distribution in our datasets. These distributions, as well as gamma and inverse gamma distributions, are special cases (or limits) of the so-called generalized inverse Gaussian distribution. For an overview of these distributions the reader is referred to Jørgensen (1982).

reasonable to apply standard asymptotic results in statistical inference.

We apply the models described above to the daily price ranges of the Hang Seng Index (HSI) from the time period from 11th of October 1989 to 24th of February 2006. The data are summarized in Table 7, where summary statistics for the estimation and forecasting periods are presented. Figure 6 reveals the extremely persistent autocorrelation structure of both original and squared series.

Tables 7: Summary statistics for estimation period and forecasting period for the Hang Seng Index.

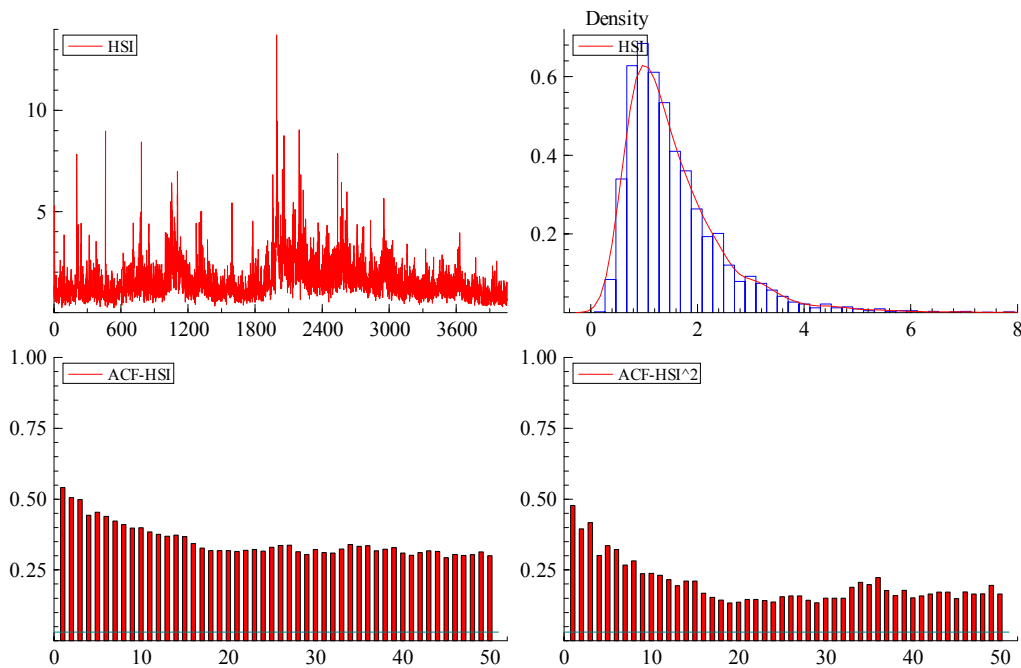|	Estimation period	Forecasting period
Observations	2054	2000
Mean	1.530	1.635
Maximum	13.724	9.035
Minimum	0.243	0.285
Standard deviation	1.051	0.949
Skewness	3.251	2.020
Kurtosis	21.774	11.991

Figure 6. Daily logarithmic price ranges of the Hang Seng Index, actual values (upper left), histogram (upper right), autocorrelation function of actual values (lower left) and squared actual values (lower right).

We employ the first 2054 observations to the estimation of the models and the 2000 last observations are saved for forecasting. Estimation results for the one component CARR model based on the gamma distribution are presented in the first column of Table 8.

Table 8: Estimation results for CARR models.

	CARR-G	MCARR-G	CARR-IG	MCARR-IG
π		0.907 (0.038)		0.551 (0.167)
γ_1	5.366 (0.163)	7.138 (0.386)	5.799 (0.176)	8.013 (1.091)
ω_1	0.034 (0.014)	0.023 (0.012)	0.025 (0.012)	0.010 (0.009)
α_{11}	0.198 (0.015)	0.188 (0.015)	0.215 (0.015)	0.142 (0.033)
α_{13}	-0.078 (0.031)	-0.098 (0.030)	-0.115 (0.030)	-0.090 (0.036)
β_1	0.858 (0.039)	0.887 (0.039)	0.884 (0.036)	0.944 (0.026)
γ_2		3.496 (0.580)		4.959 (0.460)
ω_2		0.153 (0.156)		0.070 (0.038)
α_{21}		0.299 (0.097)		0.319 (0.038)
β_{21}		0.759 (0.102)		0.621 (0.085)
$\hat{l}$	-906.20	-871.21	-863.50	-857.30
AIC	1822.40	1760.42	1737.00	1732.60
BIC	1850.53	1811.06	1765.13	1783.24
χ^2	89.50	51.55	28.05	32.56

χ^2 denotes the (Pearson) chi-square statistic for the probability integral transformed data using 25 bins. Standard errors (in parentheses) obtained from the inversed Hessian.

Based on standard errors obtained from the inversed Hessian we conclude that a model with $q = 3$ and $p = 1$ provides the best fit with the restriction $\alpha_{12} = 0$. However, the inadequacy of the model is illustrated by the upper left panel of Figure 7, where the distribution of the probability integral transformed (PIT) residuals is illustrated. The PIT residuals are calculated as

$$z_t = \int_{-\infty}^{x_t} f_{t-1}(u)du, \ t = 1, 2, ..., T. \tag{24}$$

For the correct model the PIT residuals are (approximately) independently and

uniformly distributed. Clearly this is not the case for the estimated CARR-G model, as also depicted by the respective χ^2 statistic[35] (last line of the first column of Table 8). According to upper panel of Figure 8 there is only little autocorrelation in the PIT residuals (and squared PIT residuals) of the single component model.
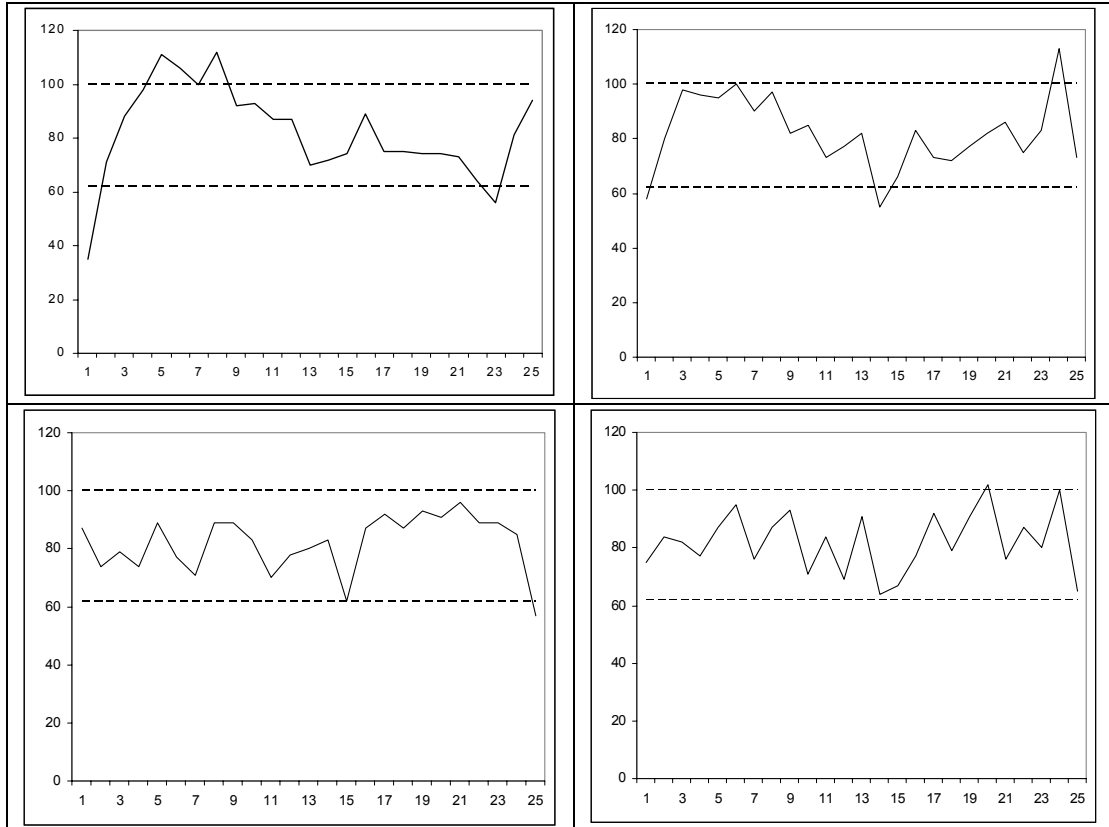
Figure 7. Plot of the distribution of the PIT data for CARR-G (upper left panel), MCARR-G (upper right), CARR-IG (lower left) and MCARR-IG (lower right) models for HSI (critical values are 62 (lower bound) and 100 (upper bound)).

---

[35]The statistic is based on the histogram of the PIT residuals and given by $\sum_{i=1}^{m}(T_i - T/m)/(T/m)$, where m is the number of bins in the histogram (we used 25 bins) and T_i is the number of observations in the bin.
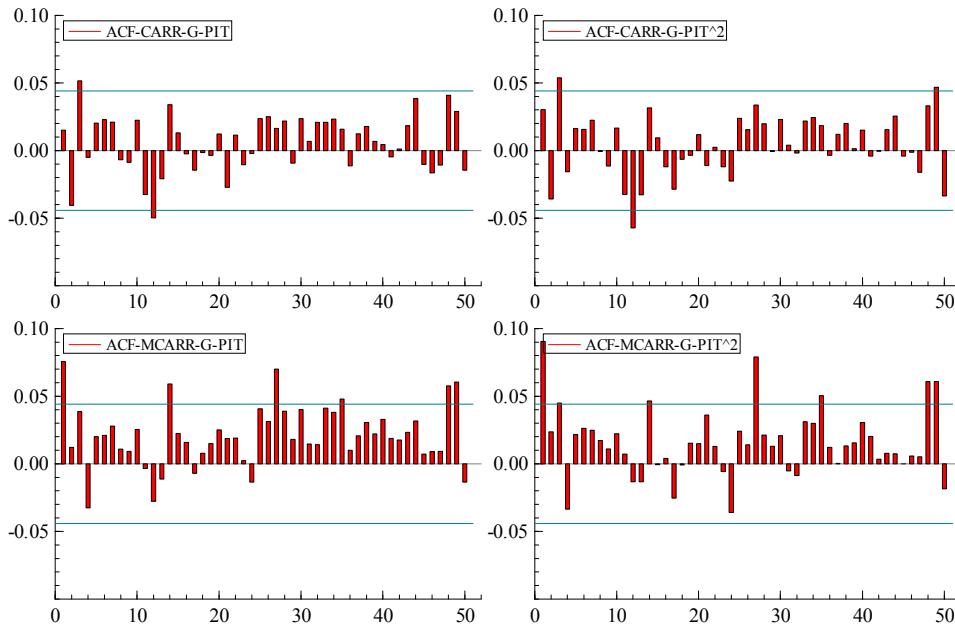
Figure 8. Autocorrelation functions of the PIT data and squared PIT data for CARR-G (upper panel)
and MCARR-G (lower panel) models for HSI.

The estimation results for the MCARR-G model are presented in the second column of Table 8. In the first component the conditional mean specification is of the order (3,1), whereas in the second component the mean equation is of the order (1,1). The estimated weight of the first component equals 0.907. Estimates for the parameters of the first component are very similar to those obtained for the single component model excluding the estimate for the shape parameter of the respective gamma distribution γ_1. This parameter is increased in the mixture model. In the second component we obtain a relatively high estimate for the constant ω_2. Estimates for the other parameters also seem to differ from those obtained in the first component. The fit of the model is much-improved according to the standard model selection criteria AIC and BIC (Table 8) likewise according to PIT data (upper right panel in Figure 7)[36]. The lower panel of Figure 8 reveals minor autocorrelation in the PIT data as well as in the squared PIT data.

While the fit is considerably improved after expanding the model to two com-

---

[36]Because the mixture model has two error terms, we refer to probability integral transformed residuals as PIT data.

ponent model, Figures 7 and 8 indicate that there could still be some misspecification. The distribution of the PIT data especially seems to differ from the uniform distribution in the first six and the last two bins. As discussed at the beginning of this subchapter, we considered several alternative distributions, all of which are subclasses of so-called *GIG* distributions. Of these distributions the inverse gamma distribution proved best for our applications. The estimation results for the single component CARR-IG model are presented in the third column of Table 8. Estimates for the parameters are hardly changed from the gamma model, but the AIC and BIC figures are substantially reduced. The lower left panel of Figure 7 shows that the distribution of the PIT residuals seems to be very close to uniform distribution. In addition, the mixture CARR-IG model was fitted (last column of Table 8). We first observe that the weight of the first component is substantially reduced from the estimate obtained for the MCARR-G model. Estimates for the parameters of the conditional mean in the first component appear to match those of the MCARR-G model. In the second component the estimates have changed slightly. According to the AIC statistic, this model is the best of the models presented here. This is also supported by the lower right panel of Figure 7, where the distribution of the PIT data is presented for the MCARR-IG model. However, the BIC and χ^2 statistics suggest that the more compact CARR-IG model is sufficient. According to Figure 9, the change of the distribution seems to have no effect on the autocorrelation structure of the PIT data and the squared PIT data when compared to the CARR-G model.

As estimated models are clearly (weakly) stationary, it is of interest to see whether their autocorrelation structure corresponds to that of the data. The autocorrelation function of the estimated MCARR-IG model against the autocorrelation function calculated from the data is presented in Figure 10. It reveals that the autocorrelation function of the mixture model decays very similarly to the sample autocorrelation function calculated from the Hang Seng Index itself. It is concluded from Figure 10 that the autocorrelation of the data is well captured by the MCARR-
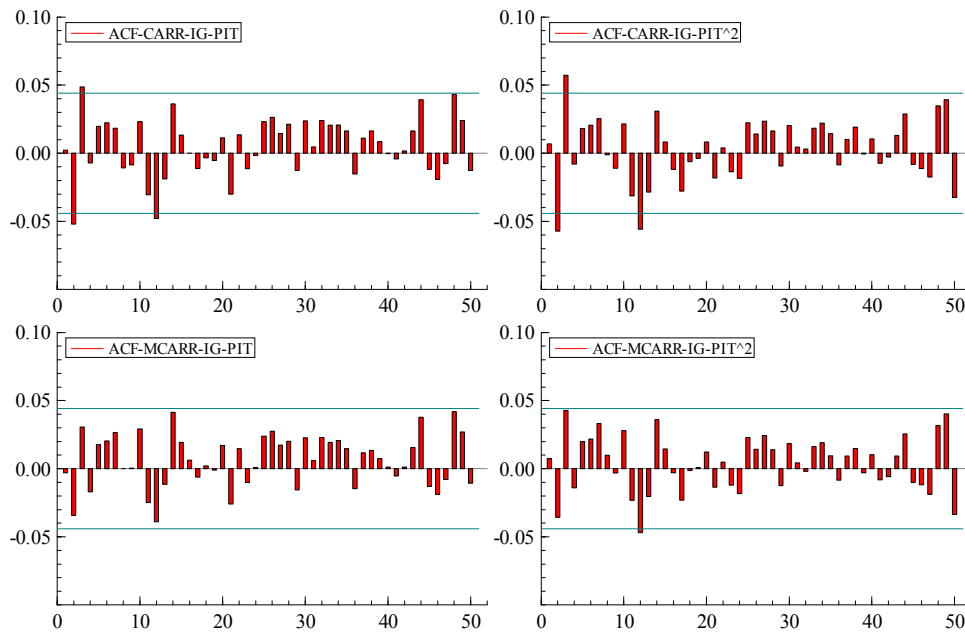
IG model.

Figure 9. Autocorrelation functions of the PIT data and squared PIT data for CARR-IG (upper panel) and MCARR-IG (lower panel) models for HSI.
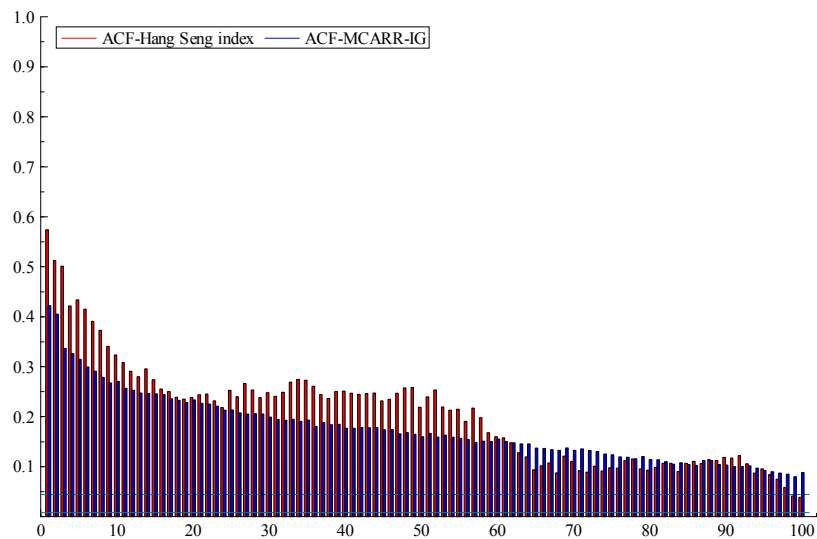

Figure 10. Autocorrelation function of actual data and autocorrelation function implied by MCARR-IG model for HSI.

We also test the out-of-sample forecasting performance of the models considered. We therefore calculated 2000 one, five and ten-day-ahead forecasts with each model. The respective mean squared errors are reported in Table 9. For the one-day-ahead forecasts (upper panel of Table 9) we find a very similar forecasting performance for the CARR-G, CARR-IG and MCARR-IG models. The differences in forecasting performance were tested using the so-called Diebold-Mariano test (see Lanne 2006 and Diebold and Mariano 1995). The nullhypothesis of the test is that the model has equal forecasting performance with the CARR-G model. According to this test there are no differences in one-day-ahead forecasting performance of the models considered.

For five-step-ahead forecasts (middle panel of Table 9) the performance of the mixture models seems to be improved by the performance of the non-mixture models. According to Diebold-Mariano tests the differences in forecasting performance are again insignificant.

Finally, for ten-day-ahead forecasts (lower panel of Table 9) the CARR-IG and MCARR-IG models seem to perform better than the CARR-G and MCARR-G models. The differences in MSEs are again insignificant but the results suggest that the choice of the conditional distribution has an impact on the forecasting performance of the model.

Table 9: Out-of-sample forecast evaluation: the Diebold-Mariano test.

Model	One-Day-Ahead Forecast		
	MSE	D-M Statistic	p-value
CARR-G	0.526		
MCARR-G	0.542	0.675	0.500
CARR-IG	0.527	0.382	0.703
MCARR-IG	0.525	0.438	0.661

Model	Five-Day-Ahead Forecast		
	MSE	D-M Statistic	p-value
CARR-G	0.557		
MCARR-G	0.569	0.959	0.338
CARR-IG	0.552	0.811	0.418
MCARR-IG	0.561	1.011	0.312

Model	Ten-Day-Ahead Forecast		
	MSE	D-M Statistic	p-value
CARR-G	0.582		
MCARR-G	0.581	0.964	0.335
CARR-IG	0.573	1.074	0.283
MCARR-IG	0.569	1.036	0.300

3.6   Discussion

This chapter discussed mixture multiplicative error models. More specifically, we considered the representation of the model as a first order random coefficient autoregressive model. This representation allowed us to express and derive conditions for the strict and weak stationarity of the model. The latter condition is needed in order to guarantee the existence of the second moments of the process, which were needed in order to derive the autocorrelation of the model.

Although the daily price range was recognized as an efficient estimator of the volatility relatively early e.g. by Parkinson (1980), the modeling of this variable has not been paid extensive attention. A multiplicative error model by Engle (2002b) was applied to the range data by Chou (2005). As an extension of the MEM model, we propose the mixture MEM model by Lanne (2006) for the daily price range of

the Hang Seng Index. The mixture structure allows flexibility in both conditional distribution and the mean dynamics.

Even if the existing literature suggests various different conditional distributions (such as exponential, gamma and Weibull) for CARR models, we find that even with the flexible MCARR specification the fit of the model can be further improved by employing the inverse gamma distribution. The implied autocorrelation function of the MCARR-IG model was found to match the autocorrelation function of the data quite well.

In addition, the out-of-sample forecast prediction performance of the models was examined. We considered one, five and ten day-a-head forecasts. While no significant differences between models were found, the forecasting performance of the models proposed in the thesis (compared to the 'standard' CARR model) was found to increase as the forecasting horizon increased.

For a topic of future research a natural direction would be to consider model specifications that allow for time varying mixture probabilities. As mentioned in the introduction, models of this kind have already been proposed in the literature at least for implied volatility and duration. Some other suggestions for future work have been mentioned by Chou (2005). In the next chapter of the thesis we consider the range as a variable in a bivariate vector multiplicative error model.

References

Ahoniemi, K. & M. Lanne 2009: Joint modeling of call and put implied volatility, *International Journal of Forecasting*, 25, 239-258.

Alexander, C. & E. Lazar 2006: Normal mixture GARCH(1,1): Applications to exchange rate modelling, *Journal of Applied Econometrics*, 21, 307-336.

Bollerslev, T. 1986: Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 31, 301-327.

Bougerol, P. & N. Picard 1992a: Stationarity of GARCH processes and of some nonnegative time series, *Journal of Econometrics,* 52, 115-127.

Bougerol, P. & N. Picard 1992b: Strict stationarity of generalized autoregressive processes, *The Annals of Probability*, 20, 1714-1730.

Brunetti, C. & P. M. Lildholdt 2007: Time series modelling of daily log-price ranges for SF/USD and USD/GBP, *The Journal of Derivatives*, 15, 2, 39-59.

Chou, R. Y. 2005: Forecasting financial volatilities with extreme values: The conditional autoregressive range (CARR) model, *Journal of Money, Credit and Banking*, 37, 561-582.

Chou, R. Y. 2006: Modeling the asymmetry of stock movements using price ranges, *Advances in Econometrics*, 20A, 231-257.

De Luca, G. & P. Zuccolotto 2003: Finite and infinite mixtures for financial durations, *Metron*, 61, 431-455.

Diebold, F. X. & R. S. Mariano 1995: Comparing predictive accuracy, *Journal of Business & Economic Statistics*, 13, 253-263.

Engle, R. F. 2002: New frontiers for ARCH models, *Journal of Applied Econometrics*, 17, 425-446.

Engle, R. F. & J. E. Russel 1998: Autoregressive conditional duration: A new model for irregularly spaced transaction data, *Econometrica*, 66, 1127-1162.

Haas, M., S. Mittnik & M. S. Paolella 2004: Mixed normal conditional heteroskedasticity, *Journal of Financial Econometrics*, 2, 211-250.

Jørgensen, B. 1982: Statistical Properties of the Generalized Inverse Gaussian Distribution. New York: Springer-Verlag.

Lanne, M. 2006: A mixture multiplicative error model for realized volatility, *Journal of Financial Econometrics*, 4, 594-616.

Liu, J.-C. 2006: Stationarity of a Markov-switching GARCH model, *Journal of Financial Econometrics*, 4, 573-593.

Nelson, D. B. & C. Q. Cao 1992: Inequality constraints in the univariate GARCH model, *Journal of Business & Economic Statistics*, 10, 229-235.

Appendix 1

Consider the GARCH (Generalized Autoregressive Conditional Heteroskedasticity) process of Bollerslev (1986). A sequence of real random variables $\{y_t, \ t \in \mathbb{Z}\}$ is said to be a GARCH(p, q) process if

$$y_t = \sqrt{h_t} v_t,$$

where innovations v_t are identically and independently distributed with zero mean and unit variance, or $v_t \sim D(0, 1)$, and the conditional variance process satisfies

$$h_t = \omega + \sum_{i=1}^{p} \beta_i h_{t-i} + \sum_{j=1}^{q} \alpha_j y_{t-j}^2, \quad t \in \mathbb{Z}, \tag{25}$$

where ω, β_i, $1 \leq i \leq p$ and α_j, $1 \leq j \leq q$ are non-negative constants.

Now suppose that $p, q \geq 2$ and define

$$
\begin{aligned}
\tau_t &= \begin{bmatrix} \beta_1 + \alpha_1 v_t^2 & \beta_2 & \ldots & \beta_{p-1} \end{bmatrix} \in \mathbb{R}^{p-1}, \\
\xi_t &= \begin{bmatrix} v_t^2 & 0 & \ldots & 0 \end{bmatrix} \in \mathbb{R}^{p-1}, \\
\alpha &= \begin{bmatrix} \alpha_2 & \ldots & \alpha_{q-1} \end{bmatrix} \in \mathbb{R}^{p-1}.
\end{aligned}
$$

Furthermore, define the $(p + q - 1) \times (p + q - 1)$ matrix

$$
A_t = \begin{bmatrix}
\tau_t & \beta_p & \alpha & \alpha_q \\
I_{p-1} & 0 & 0 & 0 \\
\xi_t & 0 & 0 & 0 \\
0 & 0 & I_{q-2} & 0
\end{bmatrix},
$$

where I_{p-1} and I_{q-2} are identity matrices of size $p - 1$ and $q - 2$. Note that the random variables $\{v_t, \ t \in \mathbb{Z}\}$ are independent. Consequently, the random matrices $\{A_t, \ t \in \mathbb{Z}\}$ are *i.i.d.* and $E(\log^+ \|A_0\|)$ is finite because all the coefficients of A_t are integrable as the distribution of y_t has finite variance. Thus the top Lyapunov exponent γ related to the sequence $\{A_t, \ t \in \mathbb{Z}\}$ is well defined. Finally, let

$$B = \begin{bmatrix} \omega & 0 & \ldots & 0 \end{bmatrix}' \in \mathbb{R}^{p+q-1},$$

and

$$X_{t-1} = \begin{bmatrix} h_t & \ldots & h_{t-p+1} & y_{t-1}^2 & \ldots & y_{n-q+1}^2 \end{bmatrix}'.$$

As y_t is a solution of (25) if and only if X_t is a solution of

$$X_t = A_t X_{t-1} + B, \quad t \in \mathbb{Z},$$

the following condition for the strict stationarity of GARCH process is readily concluded from the result given in subchapter 3.2: When $\omega > 0$, the GARCH equation (25) has a strictly stationary solution if and only if the top Lyapunov exponent γ associated with the matrices $\{A_t, \ t \in \mathbb{Z}\}$ is strictly negative. Moreover, this stationary solution is ergodic. It is the only stationary solution when ε_t's are given.

Appendix 2

As a special case of the model discussed in subchapter 3.3 consider a two-component case with mean equations of the order $(2, 2)$. We also assume that $\beta_{i,l,k}$ are equal to zero for all $i \neq l$. The representation $X_t = A_t X_{t-1} + B$ thus reduces to

$$
\begin{bmatrix} \mu_{1,t} \\ \mu_{1,t-1} \\ \mu_{2,t} \\ \mu_{2,t-1} \\ x_{t-1} \end{bmatrix}
=
\begin{bmatrix}
\beta_{1,1} + \alpha_{1,1} I(\eta_{t-1} = 1)\varepsilon_{1,t-1} & \beta_{1,2} & \alpha_{1,1} I(\eta_{t-1} = 2)\varepsilon_{2,t-1} & 0 & \alpha_{1,2} \\
1 & 0 & 0 & 0 & 0 \\
\alpha_{2,1} I(\eta_{t-1} = 1)\varepsilon_{1,t-1} & 0 & \beta_{2,1} + \alpha_{2,1} I(\eta_{t-1} = 2)\varepsilon_{2,t-1} & \beta_{2,2} & \alpha_{2,2} \\
0 & 0 & 1 & 0 & 0 \\
I(\eta_{t-1} = 1)\varepsilon_{1,t-1} & 0 & I(\eta_{t-1} = 2)\varepsilon_{2,t-1} & 0 & 0
\end{bmatrix}
$$

$$
\times
\begin{bmatrix} \mu_{1,t-1} \\ \mu_{1,t-2} \\ \mu_{2,t-1} \\ \mu_{2,t-2} \\ x_{t-2} \end{bmatrix}
+
\begin{bmatrix} \omega_1 \\ 0 \\ \omega_2 \\ 0 \\ 0 \end{bmatrix}.
$$

Appendix 3

Series (22) converges quadratically, if

$$S_{t,N} = \sum_{k=1}^{N} A_t \ldots A_{t-k+1} B$$

satisfies the Cauchy criterion. This means that for $\forall \epsilon > 0$, $\exists N_0$ such that for every $N \geq N_0$ and for every $M > 0$

$$E \left\| S_{t,N+M} - S_{t,N} \right\|^2 = E \left\| \sum_{k=N+1}^{N+M} A_t \ldots A_{t-k} B \right\|^2 < \epsilon,$$

where $\left\| \cdot \right\|$ now denotes the usual Euclidean vector norm. Denote

$$Z_{t,k} = A_t \ldots A_{t-k+1} B,$$

so that

$$S_{t,N} = \sum_{k=1}^{N} Z_{t,k}$$

and

$$E \left\| S_{t,M+N} - S_{t,N} \right\|^2 = E \left(\sum_{k=N+1}^{N+M} Z_{t,k} \right)' \left(\sum_{l=N+1}^{N+M} Z_{t,l} \right).$$

By the triangle inequality and the Cauchy-Schwarz inequality we have

$$
\begin{aligned}
\left(\sum_{k=N+1}^{N+M} Z_{t,k} \right)' \left(\sum_{l=N+1}^{N+M} Z_{t,l} \right) &= \sum_{k=N+1}^{N+M} \sum_{l=N+1}^{N+M} Z'_{t,k} Z_{t,l} \\
&= \sum_{k=N+1}^{N+M} Z'_{t,k} Z_{t,k} + 2 \sum_{k=N+1}^{N+M} \sum_{k>l>N+1} Z'_{t,k} Z_{t,l} \\
&\leq \sum_{k=N+1}^{N+M} Z'_{t,k} Z_{t,k} + 2 \sum_{k=N+1}^{N+M} \sum_{k>l>N+1} \left| Z'_{t,k} Z_{t,l} \right| \\
&\leq \sum_{k=N+1}^{N+M} Z'_{t,k} Z_{t,k} + 2 \sum_{k=N+1}^{N+M} \sum_{k>l>N+1} \left\| Z_{t,k} \right\| \left\| Z_{t,l} \right\|,
\end{aligned}
$$

where $\|Z_{t,k}\| = (Z'_{t,k}Z_{t,k})^{1/2}$. Taking expectations yields

$$
E\left(\sum_{k=N+1}^{N+M} Z_{t,k}\right)'\left(\sum_{l=N+1}^{N+M} Z_{t,l}\right)
$$

$$
\leq \sum_{k=N+1}^{N+M} EZ'_{t,k}Z_{t,k} + 2\sum_{k=N+1}^{N+M}\sum_{k>l>N+1} E\|Z_{t,k}\|\|Z_{t,l}\|
$$

$$
\leq \sum_{k=N+1}^{N+M} EZ'_{t,k}Z_{t,k} + 2\sum_{k=N+1}^{N+M}\sum_{k>l>N+1} (E\|Z_{t,k}\|^2)^{1/2}(E\|Z_{t,l}\|^2)^{1/2}
$$

$$
= \sum_{k=N+1}^{N+M} EZ'_{t,k}Z_{t,k} + 2\sum_{k=N+1}^{N+M} \left(EZ'_{t,k}Z_{t,k}\right)^{1/2}\sum_{k>l>N+1}\left(EZ'_{t,l}Z_{t,l}\right)^{1/2}
$$

$$
\leq \sum_{k=N+1}^{N+M} EZ'_{t,k}Z_{t,k} + 2\sum_{k=N+1}^{N+M} \left(EZ'_{t,k}Z_{t,k}\right)^{1/2}\sum_{l=N+1}^{N+M}\left(EZ'_{t,l}Z_{t,l}\right)^{1/2},
$$

where the second inequality follows from the Cauchy-Schwarz inequality for random variables. Thus the Cauchy criterion holds if the series $\sum_{k=1}^{\infty} EZ'_{t,k}Z_{t,k}$ and $\sum_{k=1}^{\infty} \left(EZ'_{t,k}Z_{t,k}\right)^{1/2}$ are convergent. We first write

$$
\begin{aligned}
EZ'_{t,k}Z_{t,k} &= E(Z'_{t,k}\otimes Z'_{t,k})vec(I_{mp+q-1}) \\
&= E(B'(\Pi_{i=0}^{k-1}A_{t-i})'\otimes B'(\Pi_{i=0}^{k-1}A_{t-i})')vec(I_{mp+q-1}) \\
&= (B'\otimes B')\left(E\left(A'_{t-k+1}\otimes A'_{t-k+1}\right)\cdots E\left(A'_t\otimes A'_t\right)\right)vec(I_{mp+q-1}),
\end{aligned}
$$

where well-known properties of the Kronecker product and the vectorization operator are used. The discussion above and the independence of the random variables η_t and $\varepsilon_{i,t}$ imply that

$$
\begin{aligned}
E\left(A'_{t-j}\otimes A'_{t-j}\right) &= E\left(\sum_{i=1}^{m} I(\eta_t = i)\left(e'_{i,t-j}\alpha' + C'\right)\otimes\sum_{i=1}^{m} I(\eta_t = i)\left(e'_{i,t-j}\alpha' + C'\right)\right) \\
&= E\left(\sum_{i=1}^{m} I(\eta_t = i)\left(\left(e'_{i,t-j}\alpha' + C'\right)\otimes\left(e'_{i,t-j}\alpha' + C'\right)\right)\right) \\
&= \sum_{i=1}^{m} \pi_i E\left(\left(e'_{i,t-j}\alpha' + C'\right)\otimes\left(e'_{i,t-j}\alpha' + C'\right)\right).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&EZ'_{t,k}Z_{t,k} \\
&= (B'\otimes B')\left(\sum_{i=1}^{m}\pi_i E\left(\left(e'_{i,t-1}\alpha' + C'\right)\otimes\left(e'_{i,t-1}\alpha' + C'\right)\right)\right)^k vec(I_{mp+q-1}).
\end{aligned}
$$

To see that the series $\sum_{k=1}^{\infty} EZ'_{t,k}Z_{t,k}$ and $\sum_{k=1}^{\infty}\left(EZ'_{t,k}Z_{t,k}\right)^{1/2}$ are both conver-

gent when the spectral radius of the matrix $\sum_{i=1}^{m} \pi_i E\left(\left(e'_{i,t-1}\alpha' + C'\right) \otimes \left(e'_{i,t-1}\alpha' + C'\right)\right)$ is smaller than one, the Jordan matrix decomposition is applied.

According to the Jordan matrix decomposition,

$$
\begin{aligned}
EZ'_{t,k}Z_{t,k} &= (B' \otimes B')\left(E\left(A'_{t-j} \otimes A'_{t-j}\right)\right)^k vec(I_{mp+q-1}) \\
&= (B' \otimes B')\, T\Lambda^k T^{-1} vec(I_{mp+q-1}) \\
&= u'\Lambda^k v,
\end{aligned}
\tag{26}
$$

where $u' = (B' \otimes B')\,T$, $v = T^{-1}vec(I_{mp+q-1})$, and

$$
\Lambda^k = \begin{bmatrix}
\Lambda_1^k & 0 & \cdots & 0 \\
0 & \Lambda_2^k & & 0 \\
\vdots & & \ddots & \\
0 & 0 & & \Lambda_s^k
\end{bmatrix},
$$

where

$$
\Lambda_i^k_{(n_i \times n_i)} = \begin{bmatrix}
\lambda_i^k & \binom{k}{1}\lambda_i^{k-1} & \binom{k}{2}\lambda_i^{k-2} & \cdots & \binom{k}{n-1}\lambda_i^{k-n_i+1} \\
0 & \lambda_i^k & \binom{k}{1}\lambda_i^{k-1} & & \binom{k}{n-2}\lambda_i^{k-n_i+2} \\
0 & 0 & \lambda_i^k & & \binom{k}{n-3}\lambda_i^{k-n_i+3} \\
\vdots & & & \ddots & \vdots \\
0 & 0 & 0 & \cdots & \lambda_i^k
\end{bmatrix}
$$

denotes the k-th power of the Jordan blocks Λ_i (λ_i are the eigenvalues of the matrix $E\left(A'_{t-j} \otimes A'_{t-j}\right)$). Consider a general nonzero element of Λ_i^k, that is, $\binom{k}{j}\lambda_i^{k-j}$, $0 \leq j < n_i$. We have

$$
\left|\binom{k}{j}\lambda_i^{k-j}\right| = \left|\frac{k(k-1)\ldots(k-j+1)\lambda_i^k}{j!\lambda_i^j}\right| \leq \left|\frac{k^j\lambda_i^k}{j!\lambda_i^j}\right|, \quad (\lambda_i \neq 0).
$$

By assuming that $|\lambda_i| < 1$ for all i it suffices to see that

$$
k^j\,|\lambda_i|^k \to 0,
$$

as $k \to \infty$.

Now, let $0 \leq |\lambda_i| < \rho < 1$, and write

$$
k^j\,|\lambda_i|^k = k^j\left(\frac{|\lambda_i|}{\rho}\right)^k \rho^k.
$$

Because $0 \leq j < n_i$,

$$
k^j\left(\frac{|\lambda_i|}{\rho}\right)^k \to 0,
$$

as $k \to \infty$, and it holds that

$$k^j \left(\frac{|\lambda_i|}{\rho} \right)^k \leq C < \infty,$$

for every k. We now have

$$\left| \binom{k}{j} \lambda_i^{k-j} \right| \leq \left| \frac{k^j \lambda_i^k}{j! \lambda_i^j} \right| \leq C_1 \rho^k, \tag{27}$$

where

$$\frac{k^j \left(\frac{|\lambda_i|}{\rho} \right)^k}{j! \lambda_i^j} \leq C_1 < \infty.$$

By the triangle inequality, (26) and (27) we have

$$
\begin{aligned}
EZ'_{t,k} Z_{t,k} &= u' \Lambda^k v \\
&\leq \sum \sum |u_n| \, |v_l| \left| [\Lambda^k]_{nl} \right| \\
&\leq \left(\sum \sum |u_n| \, |v_l| \, C_1 \right) \rho^k \\
&= C_2 \rho^k,
\end{aligned}
$$

where the constant C_2 is defined in an obvious way. This enables us to conclude that

$$\left(EZ'_{t,k} Z_{t,k} \right)^{1/2} \leq \left(C_2 \rho^k \right)^{1/2} = C_2^{1/2} \rho^{k/2} = C_2^{1/2} \left(\rho^{1/2} \right)^k \to 0,$$

as $k \to \infty$.

Appendix 4

In the special case mentioned in Appendix 2 the matrix A_t becomes

$$
\begin{aligned}
A_t &= \begin{bmatrix}
\beta_{1,1} + \alpha_{1,1}I(\eta_{t-1}=1)\varepsilon_{1,t-1} & \beta_{1,2} & \alpha_{1,1}I(\eta_{t-1}=2)\varepsilon_{2,t-1} & 0 & \alpha_{1,2} \\
1 & 0 & 0 & 0 & 0 \\
\alpha_{2,1}I(\eta_{t-1}=1)\varepsilon_{1,t-1} & 0 & \beta_{2,1} + \alpha_{2,1}I(\eta_{t-1}=2)\varepsilon_{2,t-1} & \beta_{2,2} & \alpha_{2,2} \\
0 & 0 & 1 & 0 & 0 \\
I(\eta_{t-1}=1)\varepsilon_{1,t-1} & 0 & I(\eta_{t-1}=2)\varepsilon_{2,t-1} & 0 & 0
\end{bmatrix} \\
&= I(\eta_t=1)\left(\alpha e_1 + C\right) + I(\eta_t=2)\left(\alpha e_2 + C\right) \\
&= I(\eta_t=1)\left(\begin{bmatrix} \alpha_{1,1} \\ 0 \\ \alpha_{2,1} \\ 0 \\ 1 \end{bmatrix}\begin{bmatrix} \varepsilon_{1,t} & 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} \beta_{1,1} & \beta_{1,2} & 0 & 0 & \alpha_{1,2} \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_{2,1} & \beta_{2,2} & \alpha_{2,2} \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}\right) \\
+I(\eta_t&=2)\left(\begin{bmatrix} \alpha_{1,1} \\ 0 \\ \alpha_{2,1} \\ 0 \\ 1 \end{bmatrix}\begin{bmatrix} 0 & 0 & \varepsilon_{2,t} & 0 & 0 \end{bmatrix} + \begin{bmatrix} \beta_{1,1} & \beta_{1,2} & 0 & 0 & \alpha_{1,2} \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_{2,1} & \beta_{2,2} & \alpha_{2,2} \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}\right).
\end{aligned}
$$

Thus, in this special case, the model is stationary, if the spectral radius of the matrix

$$
\begin{aligned}
\sum_{i=1}^{2}\pi_i E &\left(\left(\begin{bmatrix} 0_{((2i-2)\times 1)} \\ \varepsilon_{i,t-1} \\ 0_{((6-2i)\times 1)} \end{bmatrix}\begin{bmatrix} \alpha_{1,1} & 0 & \alpha_{2,1} & 0 & 1 \end{bmatrix} + \begin{bmatrix} \beta_{1,1} & 1 & 0 & 0 & 0 \\ \beta_{1,2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_{2,1} & 1 & 0 \\ 0 & 0 & \beta_{2,2} & 0 & 0 \\ \alpha_{1,2} & 0 & \alpha_{2,2} & 0 & 0 \end{bmatrix}\right)\right. \\
&\left.\otimes \begin{bmatrix} 0_{((2i-2)\times 1)} \\ \varepsilon_{i,t-1} \\ 0_{((6-2i)\times 1)} \end{bmatrix}\begin{bmatrix} \alpha_{1,1} & 0 & \alpha_{2,1} & 0 & 1 \end{bmatrix} + \begin{bmatrix} \beta_{1,1} & 1 & 0 & 0 & 0 \\ \beta_{1,2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_{2,1} & 1 & 0 \\ 0 & 0 & \beta_{2,2} & 0 & 0 \\ \alpha_{1,2} & 0 & \alpha_{2,2} & 0 & 0 \end{bmatrix}\right),
\end{aligned}
$$

is smaller than one.

4 Modeling financial volatility measures with a VMEM model based on an asymmetric copula

The last chapter of the thesis focuses on vector multiplicative error models (VMEM). In the multivariate MEM model, correctly describing the dependence structure of the error terms is of interest. For a more specific discussion on this, see subchapter 4.2. In this thesis we follow Cipollini et al. (2006, 2007) and use copula functions to model this dependence. Copulas are covered in more detail in subchapter 4.1. In subchapter 4.2 VMEM models are specified and different copulas for these models are considered. Finally, in subchapter 4.3 we present an application to absolute returns and daily price range of the Standard & Poor's 500 data. The advantages of our specification against previously proposed models are shown through standard model selection criteria and a forecasting application to the so-called VIX index. In particular, the explanatory power of the model based forecasts is found to increase as the fit of the copula and thus that of the model increases. Subchapter 4.4 includes discussion concerning the obtained results and directions for future research.

4.1 Copulas

In this subchapter, we introduce some of the main ideas of copulas. The following discussion is mostly based on Cipollini et al. (2006). For a more comprehensive and mathematically rigorous treatment the reader is referred to Embrechts et al. (2001) or Bouyé et al. (2000). For a textbook, see Nelsen (1999).

As defined by Cipollini et al. (2006), a K-dimensional copula C is the cumulative distribution function of a continuous uniform random variable defined on the unit hypercube $[0, 1]^K$. This means that every univariate component of the random variable has a $Uniform(0, 1)$ marginal distribution but the components are not assumed to be independent. Naturally, the associated copula density c is defined as

$$c(u) = \frac{\partial^K C(u)}{\partial u_1 ... \partial u_K},$$

where $u = (u_1, ..., u_K)$.

The usefulness of copulas arises from two results. The first one is the well known fact that if a random variable $X \sim F$, where F is the cumulative distribution function of X, then $U = F(X) \sim Uniform(0, 1)$. In proportion, if $U \sim Uniform(0, 1)$ then $F^{-1}(U) \sim F$. The latter relation is used extensively in simulation methodology.

The second result is known as Sklar's theorem. According to Sklar's theorem, if F is a K-dimensional cumulative distribution function with univariate continuous marginals $F_1, ..., F_K$, there exists a *unique* K-dimensional copula C such that, for all $x \in \mathbb{R}^K$,

$$F(x_1, ..., x_K) = C(F_1(x_1), ..., F_K(x_K)). \tag{28}$$

By contrast, if C is a K-dimensional copula and $F_1, ..., F_K$ are univariate cumulative distribution functions, then the function F is a K-dimensional cumulative distribution function with marginals $F_1, ..., F_K$. For proof see Sklar (1996).

It can also be verified that if F is a cumulative distribution function with univariate marginals $F_1, ..., F_K$ and $u_i = F_i(x_i)$, $i = 1, ..., K$, then $F(F_1^{-1}(x_1), ..., F_K^{-1}(x_K))$ is a copula. This guarantees the representation in (28) which can be written as

$$C(u_1, ..., u_K) = F(F_1^{-1}(x_1), ..., F_K^{-1}(x_K)). \tag{29}$$

It may be noted that (29) enables a direct way to find copulas.

As discussed by Bouyé et al. (2000, 5), the importance of Sklar's theorem arises from the fact that it provides a way to analyze the dependence structure of multivariate distributions without studying the marginal distributions. They further discuss that in financial applications the problem is often not to use a given multivariate distribution but to find a convenient distribution to describe some stylized facts. This is of importance, because the often used multivariate Gaussian distribution, assumed for tractable calculus, may not always be appropriate. It is also worth mentioning that Cipollini et al. (2006) and Bouyé et al. (2000) point out that copulas provide a powerful tool for finance because they allow the modeling problem

to be split into two steps: the first is the identification of the marginal distributions and the second defining an appropriate copula in order to represent the dependence structure in an appropriate manner. This chapter proposes a contribution to both of these steps, for we propose both marginals and copulas not so far extensively applied in the previous literature to model financial volatility measures.

We now take a look at how copulas are constructed from known multivariate distributions. These copulas are known in the literature as implicit copulas. After a straightforward derivation one obtains from (29) that the associated copula density c given in terms of the density f and marginals $f_1, ..., f_K$ is

$$c(u_1, ..., u_K) = \frac{f(F_1^{-1}(x_1), ..., F_K^{-1}(x_K))}{f_1(F_1^{-1}(x_1)) \cdots f_K(F_K^{-1}(x_K))}. \tag{30}$$

Next recall the density function of the multivariate Normal distribution

$$f(x_1, ..., x_K) = \frac{1}{(2\pi)^{K/2}(|\Sigma|)^{1/2}} \exp\left[-\frac{1}{2}(x - \mu)'\Sigma^{-1}(x - \mu)\right],$$

where μ and Σ denote the mean and the covariance matrix of the distribution. Implicit copulas do not have a simple closed form. The normal copula is based on the standardized distribution, in which μ is the null-vector and the covariance matrix equals the correlation matrix R. For a bivariate case the Normal copula has the representation (Aas 2004)

$$C_N(u_1, u_2) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi(1 - \rho^2)^{1/2}} \exp\left\{-\frac{x^2 - 2\rho xy + y^2}{2(1 - \rho^2)}\right\} dx\, dy,$$

where ρ is the parameter of the copula (the correlation coefficient) and Φ^{-1} denotes the inverse of the cumulative distribution function of a standard Normal distribution.

From (30) it follows that the copula density for a normal copula reads as

$$
\begin{aligned}
c_N(u_1, ..., u_K) &= \frac{\frac{1}{(2\pi)^{K/2}(|R|)^{1/2}} \exp\left[-\frac{1}{2}q'R^{-1}q\right]}{\prod_{j=1}^{K} \frac{1}{(2\pi)^{1/2}} \exp\left[-\frac{1}{2}q_j^2\right]} \\
&= |R|^{-1/2} \exp\left[-\frac{1}{2}q'(R^{-1} - I_K)q\right],
\end{aligned}
$$

where $q = (q_1, ..., q_K) = (\Phi^{-1}(u_1), ..., \Phi^{-1}(u_K))$ and I_K is an identity matrix of the order K. For simulation from this copula, see Embrechts et al. (2001, 26).

From (28) it follows that the density of a K-dimensional random variable X that has a copula representation can be written as

$$
f(x_1, ..., x_K) = c(F_1(x_1), ..., F_K(x_K)) \prod_{i=1}^{K} f_i(x_i).
$$

Now we are able to express the probability density function of a multivariate random variable X that has gamma marginals (with scale parameters restricted to the inverses of shape parameters) and the dependence of the components is modeled by the normal copula:

$$
f(x_1, ..., x_K) = |R|^{-1/2} \exp\left[-\frac{1}{2}q'(R^{-1} - I_K)q\right] \prod_{i=1}^{K} \frac{\phi_i^{\phi_i}}{\Gamma(\phi_i)} x_i^{\phi_i - 1} \exp(-\phi_i x_i),
$$

where $q = (\Phi^{-1}(F_1(x_1)), ..., \Phi^{-1}(F_K(x_K)))$ and $F_i(x_i) = \Gamma(\phi_i, \phi_i x_i)$ with $\Gamma(\cdot, \cdot)$ denoting the (lower) incomplete gamma function $\Gamma(x, y) = \int_0^y t^{x-1} e^{-t} dt$.

In our applications we also consider the density where the J first components of X have inverse gamma distributions (with scale parameters set as shape parameters -1) as marginal distributions and the $K - J$ last components have gamma distributions as before (for a motivation of this formulation, see subchapters 3.5 and 4.3).

This density reads as

$$
\begin{aligned}
f(x_1, ..., x_K) \;=\;& |R|^{-1/2} \exp\left[-\frac{1}{2} x'(R^{-1} - I_K)x \right] \\
&\times \prod_{i=1}^{J} \frac{(\gamma_i - 1)^{\gamma_i}}{\Gamma(\gamma_i)} x_i^{-\gamma_i - 1} \exp\left(-\frac{\gamma_i - 1}{x_i} \right) \\
&\times \prod_{i=J+1}^{K} \frac{\phi_i^{\phi_i}}{\Gamma(\phi_i)} x_i^{\phi_i - 1} \exp(-\phi_i x_i)
\end{aligned}
$$

where $x = (\Phi^{-1}(F_1(x_1)), ..., \Phi^{-1}(F_K(x_K)))$ with $F_i(x_i) = 1 - \Gamma(\gamma_i, (\gamma_i - 1)/x_i)$, $i = 1, ..., J$, and $F_i(x_i) = \Gamma(\phi_i, \phi_i x_i)$, $i = J + 1, ..., K$.

Another well known implicit copula is the Student T copula. For a bivariate case it is given by

$$
C_T(u_1, u_2) = \int_{-\infty}^{T^{-1}(u_1;v)} \int_{-\infty}^{T^{-1}(u_2;v)} \frac{1}{2\pi(1 - \rho^2)^{1/2}} \left\{ 1 + \frac{x^2 - 2\rho xy + y^2}{v(1 - \rho^2)} \right\} dx \; dy,
$$

where $v > 2$ denotes the degrees of freedom parameter, $T^{-1}(x;v)$ is the inverse of the standard cumulative distribution function of the T distribution with zero mean and variance $v/(v - 2)$, and ρ is the correlation parameter. Similarly to the Normal copula, the copula density (now for K dimensional u) is given by

$$
c_T(u_1, ..., u_K) = \frac{\Gamma((v + K)/2)\Gamma(v/2)^{K-1}}{\Gamma((v + 1)/2)} |R|^{-1/2} \frac{(1 + q'R^{-1}q/v)^{-(v+K)/2}}{\prod_{i=1}^{K}(1 + q_i^2/v)^{-(v+1)/2}},
$$

where $q = (T^{-1}(u_1;v), ..., T^{-1}(u_K;v))$, $T(x;v)$ denotes the cumulative distribution function of the Student T distribution computed at x, and $\Gamma(\cdot)$ is the gamma function.

As already discussed, the problem in many applications is finding a copula that is capable of describing the data at hand. The problem with implicit copulas, such as the mentioned Normal and T copulas, is that they have a decidedly symmetric behavior. This is also true for a larger class of implicit copulas such as elliptical copulas (see Cipollini et al. 2007, 8-9 for a discussion). What is meant by symmetry

is best described by a discussion about bivariate copulas taken from Hurd et al. (2007, 15-17). The authors divide the asymmetry of (bivariate) copulas into two parts, the first being the asymmetry along the 45 degree line and the second the asymmetry across the 45 degree line. The first asymmetry is expressed in terms of a bivariate copula density as

$$c(u_1, u_2) \neq c(1 - u_2, 1 - u_1).$$

In this type of asymmetry the joint negative events are more dependent than positive events or vice versa. There are several copulas, such as the Gumbel copula or the Clayton copula, that can model this asymmetry. These copulas are examples of so-called Archimedean copulas, which will be discussed after describing the other type of asymmetry, the asymmetry across the 45 degree line. In terms of bivariate copulas this asymmetry is written as

$$c(u_1, u_2) \neq c(u_2, u_1).$$

In this case the copula is not interchangeable, that is, $C(u_1, u_2) \neq C(u_2, u_1)$. This type of asymmetry has been discussed very little in the existing literature, the only reference known to the present author being Mari and Monbet (2004), where asymmetric extensions to the Clayton and Gumbel copulas are developed.

We now define the already mentioned Archimedean copulas. According to Bouyé et al. (2000, 18) the Archimedean copulas are defined by the equation

$$C(u_1, ..., u_K) = \begin{cases} \varphi^{-1}(\varphi(u_1) + ... + \varphi(u_K)) \text{ if } \sum_{i=1}^{K} \varphi(u_i) \leq \varphi(0) \\ 0 \text{ otherwise} \end{cases}$$

where the so-called the generator of copula $\varphi(u_i)$ has the properties $\varphi(1) = 0$, $\varphi'(u_i) < 0$ and $\varphi''(u_i) > 0$ for all $0 \leq u_i \leq 1$[37]. For example, for $\varphi(u_i) = (-\ln(u_i))^a$, $a \geq 1$

---

[37]The inverse φ^{-1} is assumed to be *completely monotonic* on $[0, \infty)$, i.e. $(-1)^k \frac{d^k}{dt^k} \varphi^{-1}(t) \geq 0$ for any $t \in (0, \infty)$ and $k \in \mathbb{N}_0$.

and $K = 2$ we obtain the bivariate Gumbel copula

$$C_G(u_1, u_2) = \exp(-(\widetilde{u}_1^a + \widetilde{u}_2^a)^{\frac{1}{a}}),$$

where $\widetilde{u}_i = -\ln(u_i)$. For a discussion on this and several other Archimedean copulas, see, for example, Venter (2001).

The (bivariate) Gumbel copula has been generalized to an asymmetric Gumbel copula by Mari and Monbet (2004). This copula is given by

$$C_{AG}(u_1, u_2) = u_1^A u_2^B \exp(-a_0 H^{\frac{1}{a_3}}),$$

where

$$A = \frac{a_1}{a_0 + a_1}, \ B = \frac{a_2}{a_0 + a_2}, \ H = \left(\frac{\widetilde{u}_1}{F}\right)^{a_3} + \left(\frac{\widetilde{u}_2}{G}\right)^{a_3}$$

and

$$F = a_0 + a_1, \ G = a_0 + a_2.$$

The parameters satisfy conditions $a_i \geq 0$, $i = 0, 1, 2$ and $a_3 \geq 1$. The standard bivariate Gumbel copula is obtained when $a_0 = 1$ and $a_1 = a_2 = 0$. The density of the asymmetric Gumbel copula is

$$c_{AG}(u_1, u_2) = \frac{C_{AG}(u_1, u_2)}{u_1 u_2}(T_1 T_2 + T_3),$$

where

$$
\begin{aligned}
T_1 &= A + \frac{a_0}{F}\left(\frac{\widetilde{u}_1}{F}\right)^{a_3-1} H^{\frac{1}{a_3}-1} \\
T_2 &= B + \frac{a_0}{G}\left(\frac{\widetilde{u}_2}{G}\right)^{a_3-1} H^{\frac{1}{a_3}-1} \\
T_3 &= \frac{a_0(a_3-1)}{FG}\left(\frac{\widetilde{u}_1\widetilde{u}_2}{FG}\right)^{a_3-1} H^{\frac{1}{a_3}-2}.
\end{aligned}
$$

As we shall see in subchapter 4.3, in our applications this copula proves to be very useful, outlining the fact that symmetric copulas, such as elliptical copulas, may not be sufficient for financial data, especially volatility measures. We also present evidence that the copula employed should be able to account for asymmetries both across and along the 45 degree line. Thus copulas such as the Clayton and Gumbel copulas may not be sufficient but a more flexible specification should be used.

As discussed by Ahoniemi and Lanne (2009), the methodology presented in Diebold et al. (1999) can be applied to check the goodness-of-fit of a multivariate time series model. This methodology includes the checking of probability integral transformed (PIT) marginals as well as checking the conditional PIT data. For bivariate multivariate models based on copulas these conditional probabilities can be calculated by using the relation (see Aas et al. 2009 for higher dimensions)

$$F(x_1|x_2) = \frac{\partial C_{u_1 u_2}(F_{x_1}(x_1), F_{x_2}(x_2))}{\partial F_{x_2}(x_2)}.$$

When the marginals are uniform

$$F(u_1|u_2) = \frac{\partial C_{u_1 u_2}(u_1, u_2)}{\partial u_2}.$$

Explicit expressions for the Normal and T copulas are derived by Aas et al. (2009) the first one being

$$F_N(u_1|u_2) = \Phi\left(\frac{\Phi^{-1}(u_1) - \rho \Phi^{-1}(u_2)}{\sqrt{1-\rho^2}}\right),$$

and the second

$$F_T(u_1|u_2) = T\left(\frac{T^{-1}(u_1; v) - \rho T^{-1}(u_2; v)}{\sqrt{\frac{\left(v + (T^{-1}(u_2; v))^2\right)(1-\rho^2)}{v+1}}}; v+1\right).$$

For the asymmetric Gumbel copula we obtain after straightforward derivation

$$F_{AG}(u_1|u_2) = \frac{1}{u_2} C_{AG}(u_1, u_2) T_2.$$

4.2 Vector multiplicative error models

In this subchapter we discuss the vector multiplicative error model (VMEM) specified by Engle (2002). Let x_t be a (K-dimensional) time series generated by

$$x_t = \mu_t \odot \varepsilon_t,$$

where μ_t is a $K \times 1$ vector of conditional means, $\odot$ is the element by element (Hadamard) product, and $\varepsilon_t|\Omega_{t-1} \sim D^+(1, \Sigma)$, that is, conditional on the information set $\Omega_{t-1} = \{x_{t-1}, x_{t-2}, ...\}$, ε_t ($K \times 1$) has a non-negative distribution with ($K \times 1$) vector of ones as mean and a general positive definite ($K \times K$) variance-covariance matrix Σ. Furthermore, the innovations ε_t are assumed to be independently and identically distributed. The property for $\varepsilon_t|\Omega_{t-1} \sim D(1, \Sigma)$ guarantees that

$$
\begin{aligned}
E(x_t|\Omega_{t-1}) &= \mu_t \\
Var(x_t|\Omega_{t-1}) &= \mu_t \mu_t' \odot \Sigma = \text{diag}(\mu_t)\Sigma\text{diag}(\mu_t),
\end{aligned}
$$

where the latter is a positive definite matrix when the components of μ_t are positive.

The equation for the conditional expectation of x_t is specified by Cipollini et al. (2006, 2007) as (the so-called 'base specification')

$$\mu_t = \omega + \alpha x_{t-1} + \beta \mu_{t-1},$$

where the parameter matrices ω, α and β have dimensions $K \times 1$, $K \times K$ and $K \times K$ respectively. This is further generalized by

$$\mu_t = \omega + \alpha x_{t-1} + \beta \mu_{t-1} + \delta x_{t-1}^{(-)}, \tag{31}$$

where the $K \times 1$ vector $x_{t-1}^{(-)}$ contains components of x_{t-1} multiplied by a variable that takes a value of one when some variable (usually previous day's return r_{t-1}) is negative and otherwise zero, and δ is a $K \times K$ parameter matrix. In order to guarantee the non-negativity of the mean process μ_t, the parameters are assumed to satisfy the conditions $\omega \geq 0$, $\alpha_{ij} \geq 0$, and $\beta_{ij} \geq 0$, $\alpha_{ij} + \delta_{ij} \geq 0$ for all $i, j = 1, ..., K$. A sufficient condition for the (weak) stationarity of the VMEM model defined by Equation (31) is given by Cipollini et al. (2007, A.1). According to their result the model is stationary in mean if all characteristic roots of $A = \alpha + \beta + \delta/2$ are smaller than 1 in modulus.

As mentioned in the previous chapter (subchapter 3.2), univariate MEM models nest the squared GARCH model of Bollerslev (1986), the ACD model by Engle and Russel (1998) as well as the CARR model of Chou (2005). Cipollini et al. (2007) note that VMEM models can be applied in several contexts such as volatility forecasting, and modeling volatility spillovers, order execution dynamics and trades, duration, volume and volatility dynamics.

In this subchapter we specify the distribution of ε_t by using copulas introduced in the previous subchapter. Other possibilities would include, for example, multivariate gamma distributions. Cipollini et al. (2007) mention the multivariate gamma distribution of Cheriyan and Ramabhadran as a possible candidate. They also point out that this distribution admits only positive correlations between components. Its probability density function is also very complicated and enforces limiting constraints for the shape parameters of the marginals. Recently, Ahoniemi and Lanne (2009) proposed a bivariate gamma distribution of Nagao and Kadoa (for a reference see Yue et al. 2001). In their model the shape parameters of the marginals, however, are restricted to be equal. This assumption is likely to be too restrictive for our applications. In fact, instead of allowing for different shape parameters we wish to be even more general and use marginals belonging to different classes of distributions. More specifically, in our bivariate example we model the first marginal with an inverse gamma distribution whereas the second has a gamma distribution.

Now, the conditional distribution of ε_t in our VMEM model based on a copula

is given simply[38] by

$$f(\varepsilon_t|\Omega_{t-1}) = c(F_1(\varepsilon_{1,t}), ..., F_K(\varepsilon_{K,t})) \prod_{i=1}^{K} f_i(\varepsilon_{i,t}),$$

where c is the density of the applied copula and f_i and F_i are the probability density function and cumulative distribution function of the i-th marginal. A straightforward application of the change of variables theorem shows that the conditional distribution of x_t is given by

$$f(x_t|\Omega_{t-1}) = c(F_1(x_{1,t}/\mu_{1,t}), ..., F_K(x_{K,t}/\mu_{K,t})) \prod_{i=1}^{K} f_i(x_{i,t}/\mu_{i,t})/\mu_{i,t},$$

and the log-likelihood of a sample of size T is

$$
\begin{aligned}
l &= \sum_{t=1}^{T} l_t = \sum_{t=1}^{T} \ln f(x_t|\Omega_{t-1}) \\
&= \sum_{t=1}^{T} \ln c(F_1(x_{1,t}/\mu_{1,t}), ..., F_K(x_{K,t}/\mu_{K,t})) \\
&\quad + \sum_{t=1}^{T}\sum_{i=1}^{K} \left[\ln f_i(x_{i,t}/\mu_{i,t}) - \ln \mu_{i,t} \right].
\end{aligned}
$$

Generally the log-likelihood function can be optimized directly using numerical methods. However, as pointed out by Cipollini et al. (2006, 2007), for some copulas their parameters can be obtained by some other method such as the method of moments, after which these estimates can be plugged into the likelihood function, and thus obtain a pseudo maximum likelihood function.

Assuming the Normal copula together with the first J components having inverse gamma distributions and the last $K-J$ last components having gamma distributions

---

[38]For the sake of simplicity, the dependence of both the copula density c and the marginal densities f_i on their parameters are omitted from the notation.

we obtain

$$
\begin{aligned}
l_t &= \frac{1}{2}\ln\left|R^{-1}\right| - \frac{1}{2}q_t' R^{-1} q_t + \frac{1}{2}q_t' q_t \\
&\quad + \sum_{i=1}^{J} \gamma_i \ln((\gamma_i - 1)\mu_{i,t}) - (\gamma_i + 1)\ln(x_{i,t}) - (\gamma_i - 1)\mu_{i,t}/x_{i,t} - \ln\Gamma(\gamma_i) \\
&\quad + \sum_{i=J+1}^{K} \phi_i \ln\phi_i - \ln\Gamma(\phi_i) - \ln x_{i,t} + \phi_i\left(\ln x_{i,t} - \ln\mu_{i,t} - \frac{x_{i,t}}{\mu_{i,t}}\right),
\end{aligned}
$$

with $q_t = (q_{1,t}, ..., q_{K,t}) = (\Phi^{-1}(F_1(\varepsilon_{1,t})), ..., \Phi^{-1}(F_K(\varepsilon_{K,t})))$. For this model Cipollini et al. (2006, 2007) derived the concentrated log-likelihood function obtained by first solving for R:

$$
lc = -\frac{T}{2}\left[\ln|q'q| - \sum_{i=1}^{K}\ln(q_{\cdot i}' q_{\cdot i})\right] + \text{(marginals contribution)},
$$

where $q = (q_1', ..., q_T')$ is a $T \times K$ matrix and $q_{\cdot i}$ denotes the i-th column of matrix q. In our applications we did not find any difference whether the 'original' or concentrated likelihood function was applied because both the value of the log-likelihood function and the implied estimate of the correlation coefficient remained the same.

For the T copula with the marginals as before we have

$$
\begin{aligned}
l_t &= \ln(\Gamma((v+K)/2)) + \ln(\Gamma(v/2)^{K-1}) - \ln(\Gamma((v+1)/2)) - \frac{1}{2}\ln|R| \\
&\quad - \frac{v+K}{2}\ln(1 + q_t' R^{-1} q_t/v) + \frac{v+1}{2}\sum_{i=1}^{K}\ln(1 + q_{i,t}^2/v) \\
&\quad + \sum_{i=1}^{J} \gamma_i \ln((\gamma_i - 1)\mu_{i,t}) - (\gamma_i + 1)\ln(x_{i,t}) - (\gamma_i - 1)\mu_{i,t}/x_{i,t} - \ln\Gamma(\gamma_i) \\
&\quad + \sum_{i=J+1}^{K} \phi_i \ln\phi_i - \ln\Gamma(\phi_i) - \ln x_{i,t} + \phi_i\left(\ln x_{i,t} - \ln\mu_{i,t} - \frac{x_{i,t}}{\mu_{i,t}}\right),
\end{aligned}
$$

and for the (bivariate) asymmetric Gumbel copula

$$
\begin{aligned}
l_t \;=\;& \ln C_{AG}(u_{1,t}, u_{2,t}) + \ln(T_{1,t}T_{2,t} + T_{3,t}) - \ln(u_{1,t}u_{2,t}) \\
&+ \gamma_1 \ln((\gamma_1 - 1)\mu_{1,t}) - (\gamma_1 + 1)\ln(x_{1,t}) - (\gamma_1 - 1)\mu_{1,t}/x_{1,t} - \ln\Gamma(\gamma_1) \\
&+ \phi_2 \ln \phi_2 - \ln\Gamma(\phi_2) - \ln x_{2,t} + \phi_2\left(\ln x_{2,t} - \ln\mu_{2,t} - \frac{x_{2,t}}{\mu_{2,t}}\right),
\end{aligned}
$$

with $u_t = (u_{1,t}, u_{2,t}) = (F_1(\varepsilon_{1,t}), F_2(\varepsilon_{2,t})) = (1 - \Gamma(\gamma_1, (\gamma_1 - 1)/\varepsilon_{1,t}), \Gamma(\phi_2, \phi_2\varepsilon_{2,t}))$. Obviously, here the first element is assumed to have an inverse gamma distribution and the second one a gamma distribution. This setup will be applied in our subsequent empirical analysis.

4.3 Empirical analysis

We apply the models described in the previous subchapter to two volatility indicators, namely the daily price ranges and absolute returns[39] of the S&P 500 index from the time period June 28, 1988 to June 23, 2008. Formally, let P_t be the logarithmic price of the asset observed at time t, $t = 1, 2, ..., T$. The daily price range is again defined by (23) (subchapter 3.5). Absolute returns are given by

$$
A\_R_t = 100\,|P_t - P_{t-1}|\,.
$$

The data[40] are summarized in Table 10, where summary statistics for estimation and forecasting periods are presented. Figure 11 reveals highly persistent autocorrelation structures of both the original and squared series of both variables.

---

[39]Admittedly, the application would be more interesting if variables such as realized volatility, realized absolute variation or realized bi-power variation (see Cipollini et al. 2007 for definitions) were included. Due to a lack of a proper realized volatility series and limited availability of the intraday data this is left to future work.

[40]The days when absolute returns had the value of zero were simply deleted from the data. The days when the VIX index has no value were also deleted in order to have series with equal dimensions. The motivation for the latter will be found later in this section.

Table 10: Summary statistics for estimation period and forecasting period for daily price range and absolute returns of the S&P 500 index.

Variable	Estimation period		Forecasting period	
	Range	Abs_Ret	Range	Abs_Ret
Observations	3031	3031	2000	2000
Mean	1.128	0.655	1.347	0.793
Maximum	7.658	7.113	8.479	5.574
Minimum	0.177	0.001	0.239	0.001
Standard deviation	0.716	0.647	0.811	0.757
Skewness	2.644	2.611	1.850	1.870
Kurtosis	16.91	16.43	9.015	7.816

We use the first 3031 observations for the estimation of models and save the 2000 last observations for forecasting. The estimation results for the VMEM model based on the independent copula[41] are presented in the first column of Table 11. The results for the daily price range are presented in the upper part of the column and those for the absolute returns in the middle. The marginal distribution of the daily price range is modeled with the inverse gamma distribution, whereas the gamma distribution is applied for absolute returns. The former selection is based on earlier evidence from the previous chapter, suggesting that the inverse gamma distribution provides a satisfactory fit for daily price range data.

---

[41]The independent copula is simply obtained by setting the copula contribution equal to zero in the log-likelihood function.
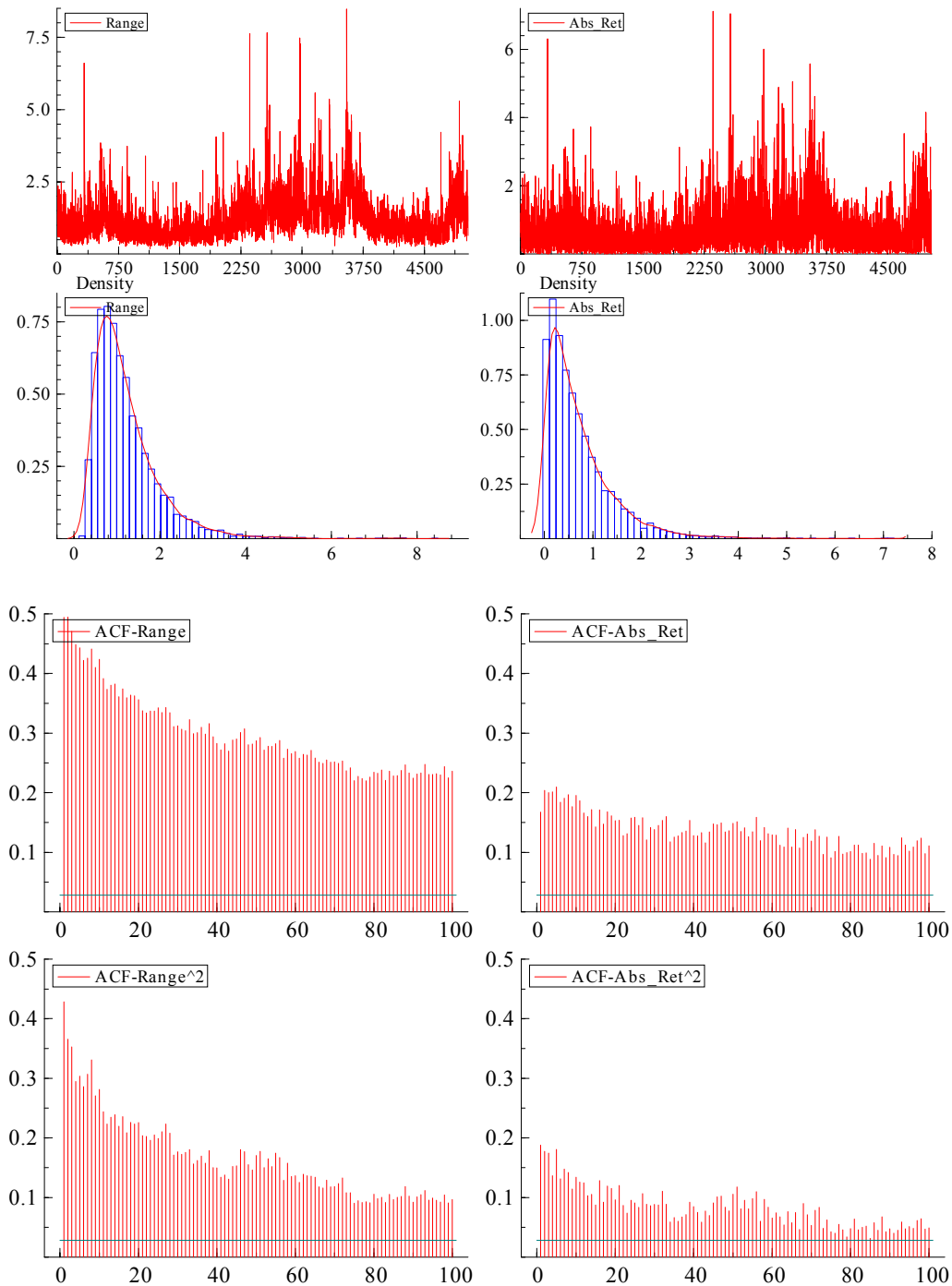
Figure 11. Daily logarithmic price ranges and absolute returns of S&P 500 index 6.27.1988 – 6.23.2008, actual values (upper panel), histograms (second panel), autocorrelation function of actual values (third panel) and squared actual values (lower panel) (solid lines for densities: kernel estimated densities, solid lines for autocorrelations: approximate 95 % confidence intervals).

Table 11: Estimation results for VMEM models.

| | | Copula | | |
	Independent	Normal	T	Asymmetric Gumbel
γ_1	5.460 (0.191)	5.467 (0.152)	5.489 (0.149)	5.342 (0.137)
ω_1	0.014 (0.004)	0.020 (0.003)	0.017 (0.003)	0.017 (0.002)
α_{11}	0.069 (0.010)	0.072 (0.008)	0.065 (0.007)	0.049 (0.006)
β_{11}	0.896 (0.010)	0.883 (0.011)	0.892 (0.007)	0.910 (0.006)
δ_{11}	0.050 (0.009)	0.055 (0.006)	0.048 (0.006)	0.055 (0.005)
ϕ_2	1.211 (0.028)	1.212 (0.028)	1.182 (0.027)	1.238 (0.025)
ω_2	0.006 (0.003)	0.009 (0.003)	0.008 (0.002)	0.010 (0.002)
α_{22}	0.011 (0.009)	0.008 (0.008)	0.009 (0.007)	0.011 (0.004)
β_{22}	0.944 (0.011)	0.933 (0.011)	0.936 (0.010)	0.933 (0.006)
δ_{21}	0.043 (0.008)	0.053 (0.007)	0.049 (0.007)	0.046 (0.004)
ρ		0.669 (0.010)	0.692 (0.013)	
v			5.223 (0.658)	
a_1				0.716 (0.047)
a_3				3.369 (0.106)
$\hat{l}$	-3248.8	-2344.7	-2290.0	-1900.7
AIC	6517.5	4711.4	4604.0	3825.4
BIC	6577.7	4777.6	4676.2	3897.6
χ^2_{Ran}	38.23	46.10	71.12	36.26
χ^2_{Abs}	40.50	31.83	36.97	34.74
$\chi^2_{Ran|Abs}$		69.38	60.67	50.55
$\chi^2_{Abs|Ran}$		692.13	742.15	163.93

Upper part: estimates for the range, middle part: estimates for the absolute returns, lower part: parameters of the copulas. Standard errors based on outer products of the first derivatives in parentheses. χ^2 denotes the (Pearson) chi-square statistic for the probability integral transformed data using 20 bins.

Based on standard errors obtained from the outer product of the first derivatives[42] we concluded that a model with diagonal parameter matrices α and β coupled with a δ matrix having nonzero elements (see Equation (31)) in the first column provided the best fit. We therefore corroborate the findings of Engle and Gallo

---

[42]These standard errors were used because of the difficulties in inverting the respective Hessian for more complicated models.

(2006) and Cipollini et al. (2006) that daily price range has explanatory power over other volatility indicators. According to Figure 12, our decision to use the inverse gamma distribution as a marginal distribution for the daily price range and gamma distribution for absolute returns proves to be reasonable. In the upper panel of this figure we observe the PIT residuals for both series. The PIT residuals are given by (24) (subchapter 3.5). For a correct model the PIT residuals are (approximately) independently and uniformly distributed. Clearly, the assumption of an approximate uniform distribution for both variables cannot be rejected. The same is depicted by the respective χ^2 statistics[43] (bottom of the first column of Table 11).
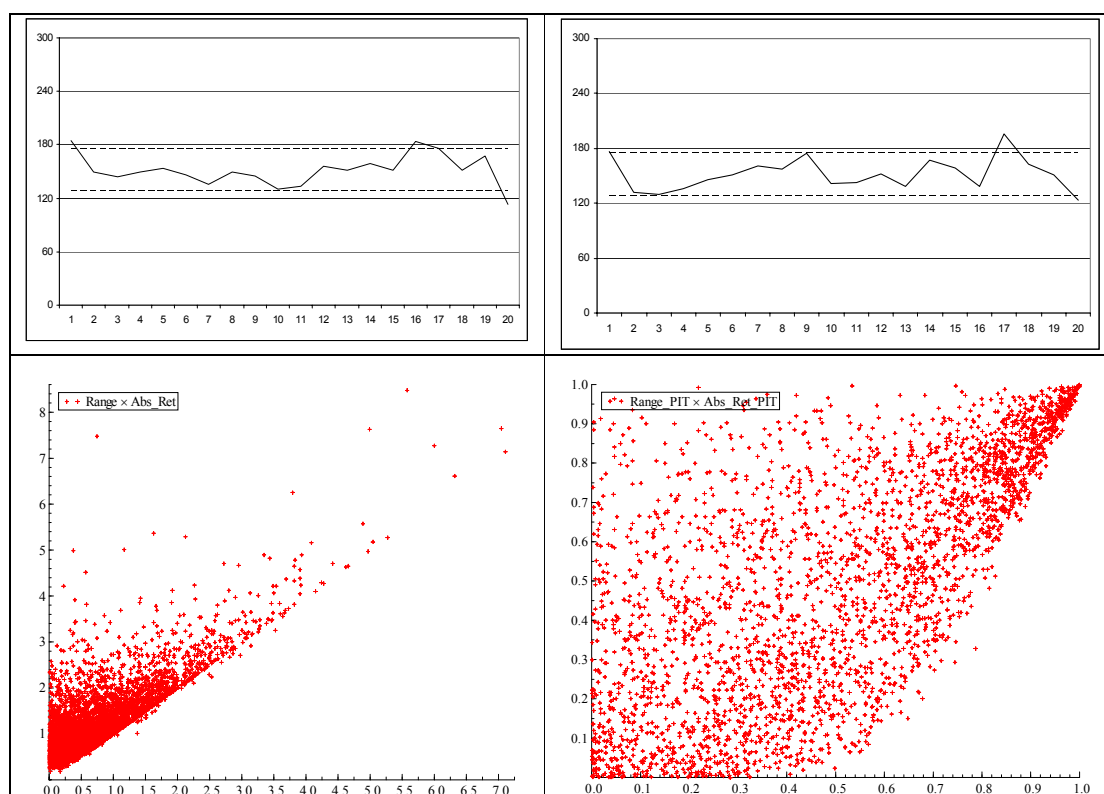

Figure 12. Plot of the distribution of the PIT data for price range (upper left panel) and absolute returns (lower right panel) for the model with independent copula (critical values (dashed lines) are 128 (lower bound) and 175 (upper bound)). A scatter plot of the actual data is presented in the lower left panel whereas a scatterplot of the PIT data for independent copula is shown in the lower right panel.

---

[43]The statistic is based on the histogram of the PIT residuals and given by $\sum_{i=1}^{m}(T_i - T/m)/(T/m)$, where m is the number of bins in the histogram (we used 20 bins) and T_i is the number of observations in the bin.

According to the scatter plot of the original observations or the PIT residuals (see lower panel of Figure 12) the variables considered cannot be considered independent and the estimated correlation of the original variables (point estimate approximately 0.7) indicates the same. Thus it is reasonable to model the dependence by using the copulas discussed in the preceding subchapters. At this point, very little is known about the usefulness of different copulas for VMEM modeling. As far as we know, the only existing studies on the problem at hand are Cipollini et al. (2006) and Cipollini et al. (2007). In these papers Normal and T copulas have been applied for modeling volatility indicators. Following this work we also estimated models using these so-called elliptical copulas.
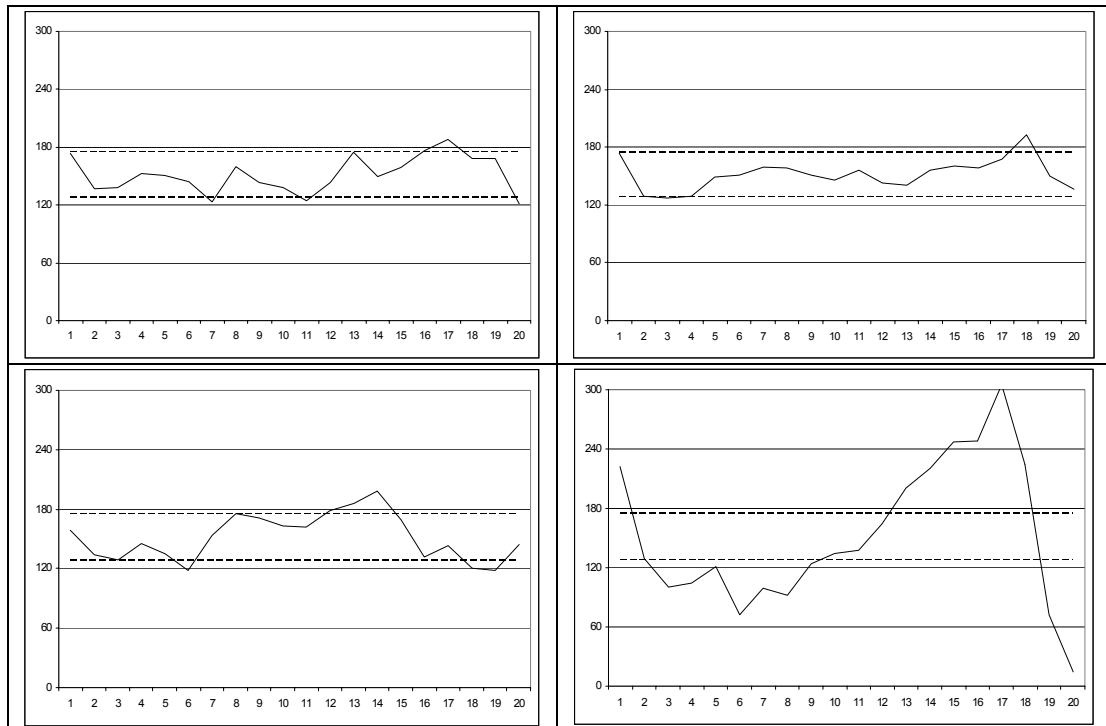

Figure 13. Plot of the distribution of the PIT data for price range (upper left panel), absolute returns (upper right), price range|absolute retuns (lower left) and absolute return|price range (lower right) for model with Normal copula (critical values are 128 (lower bound) and 175 (upper bound) (dashed lines)).

The estimation results can be found in the second and third columns of Table 11. For both models we estimated the parameters of the copulas simultaneously with the parameters in the conditional mean specification. As Table 11 shows,

the estimate for the correlation coefficient ρ is a little below 0.7 for both models. For the T copula the point estimate of the degrees of freedom parameter v is 5.2, which suggests significant deviation from the normal copula. As is known from the literature, the T copula allows us to model the tail dependence typically found in financial data. In our opinion, the estimates of the parameters of the mean equation did not change much after copulas were introduced into the model and the significance of these changes is left unanswered at this point but will be discussed in more detail later in this subchapter.
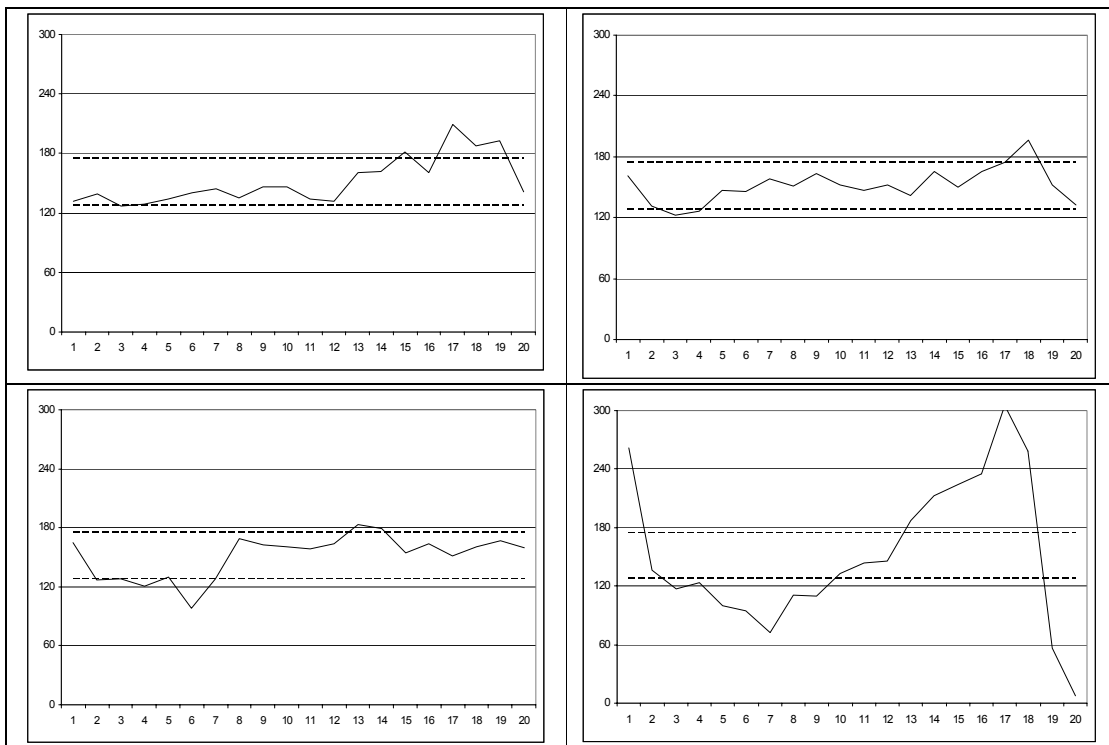

Figure 14. Plot of the distribution of the PIT data for price range (upper left panel), absolute returns (upper right), price range|absolute retuns (lower left) and absolute return|price range (lower right) for model with Student's T copula (critical values are 128 (lower bound) and 175 (upper bound) (dashed lines)).

Having the wide range of different copulas in mind, the goodness-of-fit of the copula is of special interest. We choose to follow Ahoniemi and Lanne (2009) and take a look at the PIT data using the conditional cumulative distribution functions presented in subchapter 4.2. For the normal copula model these are shown in the lower panel of Figure 13. Whereas the conditional distribution of the daily price

range given the value of the absolute return appears to be reasonably captured by the Normal copula, the converse is not the case. On the contrary, the conditional distribution of the absolute returns given the value of the price range appears to be decidedly misspecified. This is also underlined by the large value of the χ^2 statistic displayed at the bottom of the second column of Table 11.
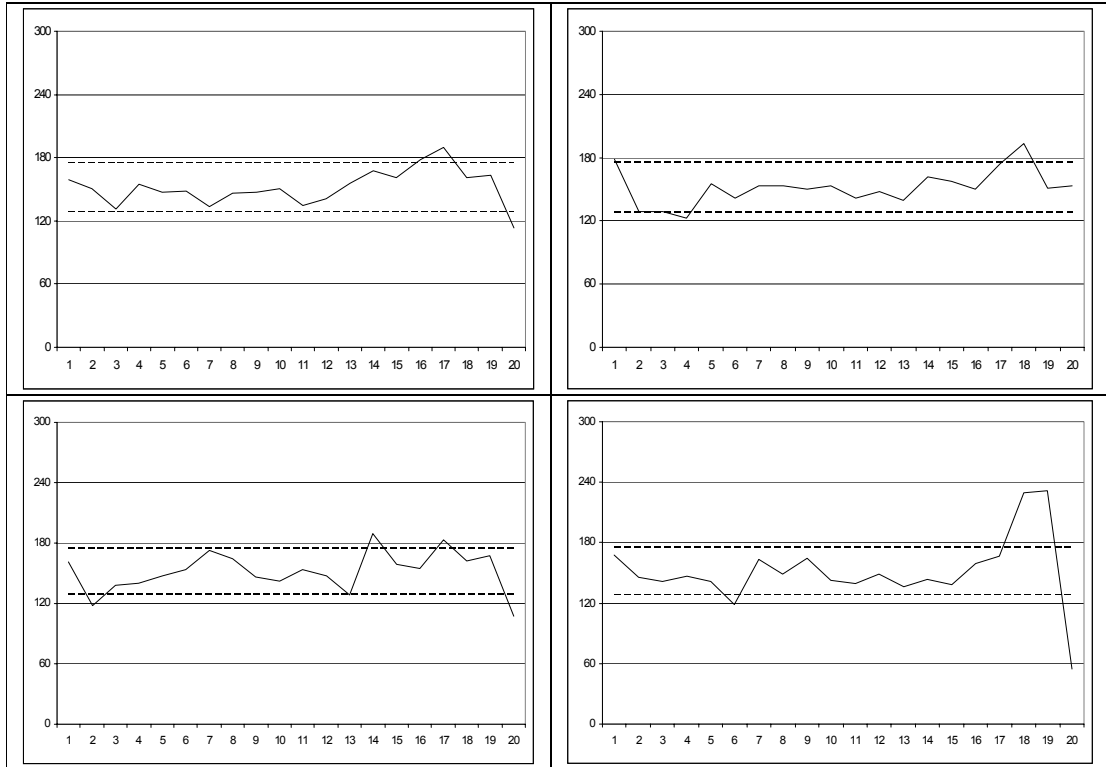

Figure 15. Plot of the distribution of the PIT data for price range (upper left panel), absolute returns (upper right), price range|absolute retuns (lower left) and absolute return|price range (lower right) for model with asymmetric Gumbel copula (critical values are 128 (lower bound) and 175 (upper bound) (dashed lines)).

Allowing for more flexibility one might expect the T copula to improve the misspecification observed for the Normal copula. However, according to the lower right panel of Figure 14 as well as the value of the χ^2 statistic in the third column of Table 11, the T copula provides no gains in fit compared to the Normal copula. The likely reason for this can be found in the lower right panel of Figure 12, where the PIT residuals of the independent copula model are plotted against each other. The figure clearly shows that the distribution of the data is asymmetric both across and along the 45 degree line. We therefore claim that the dependence of the variables used

in this application can only be adequately described by a copula able to describe both these asymmetries. We therefore estimated a VMEM model based on the asymmetric Gumbel copula of Mari and Monbet (2004).[44]

Estimation results for this model are presented in the last column of Table 11. In the estimation we found it necessary to fix the parameters a_0 and a_2 in the copula to unity and zero[45] respectively. The parameters a_1 and a_3 turned out to be accurately estimated. Estimates for the parameters of the models for the conditional mean are slightly affected by the change of the copula. However, the flexibility of the copula and thus the improved fit is revealed in the lower panel of Figure 15. It shows that especially the conditional distribution of the absolute returns is now better described (lower right panel of Figure 15). We observe a slight misspecification in the last three bins, but the improvement compared to elliptical copulas is dramatic. The same is predictably indicated by the value of the χ^2 statistics at the bottom of Table 11. The improved fit is also in line with the AIC and BIC figures in Table 11.

At this point we have demonstrated that a correct specification of the copula is relevant in VMEM modeling in order to adequately describe the dependence structure of the variables. We leave space for the argument that the specification of the copula is not relevant if one is not interested in these structures. Thus, it is of interest to see how the estimated models perform in a context that does not directly involve copulas, that is, forecasting analyzed variables.

Recently, Engle and Gallo (2006) provided evidence that (VMEM) model based volatility indicators have predictive power to a market based volatility indicator, namely the VIX index. The VIX index is essentially a volatility indicator calculated by CBOE from the S&P 500 index[46] option prices with 30 days to maturity. Engle and Gallo (2006) find that 22-day-ahead (in line with the 30-day horizon used in the

---

[44]We also tried several copulas that are able to model the asymmetry along the 45 degree line such as the Clayton copula, the Heavy Right Tail copula and the Gumbel copula. In general, we found that these copulas were able to outperform elliptical copulas but their performance was consistently exceeded by the asymmetric Gumbel copula. However, estimation results for these models are not reported in detail.

[45]These choices coincide with the values which make the asymmetric Gumbel copula to agree with the symmetric Gumbel copula.

[46]The calculation was based on S&P 100 index option prices in 1990-2003.

construction of the VIX) predictions from their VMEM model for daily price ranges, absolute returns, and realized volatility have significant predictive power when the VIX index level is modeled with an AR(1) model.

Instead of modeling the VIX index level we consider the approach of Ahoniemi (2006) and choose to model the first log-differences of the VIX index. This procedure is also emphasized by Fleming et al. (1995), who argue that the variable of interest for academics and practitioners is changes or innovations in expected volatility. They also speculate that if stock prices follow a random walk, estimation of the relationship between the stock and volatility indices may be spurious. Finally it is pointed out that because VIX levels appear to be near random walk the inference may be affected in finite samples by high autocorrelation.

The VIX index level from January 2, 1990 to June 23, 2008, its first log-differences, and their autocorrelation functions for the first 2650 observations are presented in Figure 16. Summary statistics for the first log-differences can be found in Table 12.

Table 12: Summary statistics for the VIX index log-differences.

	In-sample	Out-of-sample
Observations	2649	2000
Mean	0.000	0.000
Maximum	0.419	0.496
Minimum	-0.295	-0.300
Standard deviation	0.058	0.057
Skewness	0.662	0.591
Kurtosis	7.084	7.808

As a reference model we use an AR(1)-GARCH(1,1) specification because the plot of the first differences clearly shows volatility clustering in the series. The estimation results for the estimation period (i.e. the period overlapping with the period used in estimation of the VMEM models) are presented in the first column of Table 13 and for the forecasting period in Table 14. One can observe that the obtained parameter estimates for conditional mean and conditional variance are reasonable both in-sample and out-of-sample. We compute 22-day-ahead forecasts for both volatility indicators using the four copulas previously discussed, that is, the independent,

Normal, T and asymmetric Gumbel copulas. These forecasts are then employed as explanatory variables in the conditional mean of the AR(1)-GARCH(1,1) model.
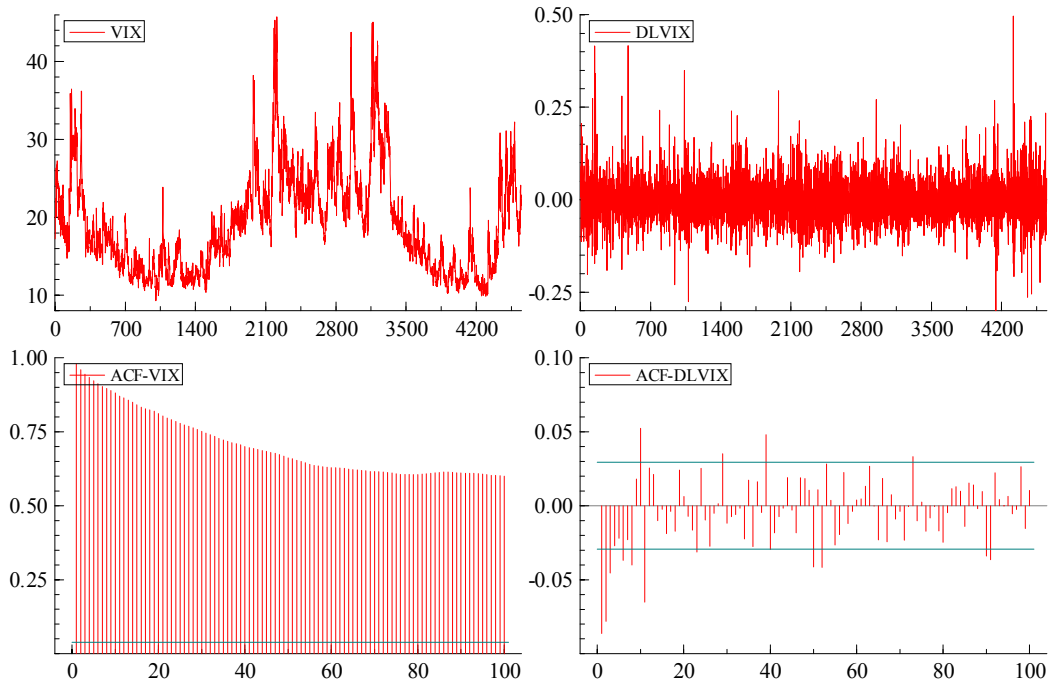

Figure 16. VIX index 1.2.1990 – 6.23.2008, actual values (upper left) and first log-differences (upper right) and their autocorrelation functions (lower panel, solid lines for autocorrelations: 95 % approximate confidence intervals).

As revealed by Table 13, the forecasts from independent, Normal and T copulas seem to have no significant explanatory power over the changes of the VIX index. By contrast, the forecasts based on the model using the asymmetric Gumbel copula prove to be significant. It is observed that the coefficient of the price range ϕ_{Ra}^* is estimated negative whereas $\phi_{A\_R}^*$, the coefficient for the absolute return, is estimated positive. This is in line with the findings of Engle and Gallo (2006). Out-of-sample forecasts from all models are significant but their absolute value is increased when the fit of the copula is increased (Table 14). Thus, forecasts from the asymmetric Gumbel copula VMEM model seem to have most predictive power to the VIX changes. This suggests that a correct specification of the copula has a significant impact on the forecasting ability of the model based volatility on market

based volatility.

Table 13: Estimation results for AR(1)-GARCH(1,1) models for the VIX log-differences, in-sample.

		Copula			
	AR(1)	Independent	Normal	T	Asymmetric Gumbel
ϕ_0^*	0.0002	-0.001	0.0008	0.003	-0.005
	(0.001)	(0.004)	(0.007)	(0.005)	(0.005)
ϕ_1^*	-0.069	-0.069	-0.069	-0.070	-0.072
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
ϕ_{Ra}^*		0.010	-0.054	-0.092	-0.228
		(0.040)	(0.056)	(0.053)	(0.066)
$\phi_{A\_R}^*$		-0.016	0.088	0.149	0.419
		(0.066)	(0.087)	(0.083)	(0.120)
ω^*	0.001	0.001	0.001	0.001	0.001
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.003)
α^*	0.096	0.096	0.095	0.094	0.093
	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)
β^*	0.748	0.747	0.748	0.749	0.751
	(0.102)	(0.102)	(0.103)	(0.104)	(0.104)
$\hat{l}$	3835.2	3835.2	3836.0	3836.9	3841.4
AIC	-7660.3	-7656.4	-7658.0	-7659.8	-7668.9
BIC	-7631.0	-7615.2	-7616.8	-7618.6	-7627.6
Out-of-sample forecasts					
MSE$\times 10^3$	3.262	3.265	3.243	3.228	3.186
MAD$\times 10^2$	4.200	4.202	4.182	4.176	4.133

Robust standard errors in parentheses. ϕ_0^* denotes the constant in the conditional mean, ϕ_1^* is the AR-parameter, ϕ_{Ra}^* and $\phi_{A\_R}^*$ are the coefficients of the range and absolute return and ω^*, α^* and β^* are parameters of the GARCH part of the models.

4.4   Discussion

The topic of this chapter has been vector multiplicative error models. In this thesis, the VMEM models considered were based on so-called copula functions. Advantages of this approach include the fact that by using copulas the modeling of the joint distribution function can be divided into two parts: first finding the appropriate marginal distributions for each variable and choosing suitable copula function to correctly describe the dependence of the marginals.

Table 14: Estimation results for AR(1)-GARCH(1,1) models for the VIX log-differences, out-of-sample.

	AR(1)	Copula			
		Independent	Normal	T	Asymmetric Gumbel
ϕ_0^*	-0.001	0.011	0.036	0.029	-0.0006
	(0.001)	(0.005)	(0.008)	(0.006)	(0.004)
ϕ_1^*	-0.091	-0.098	-0.108	-0.113	-0.126
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
ϕ_{Ra}^*		-0.197	-0.439	-0.466	-0.708
		(0.049)	(0.070)	(0.063)	(0.072)
$\phi_{A\_R}^*$		0.328	0.672	0.726	1.179
		(0.080)	(0.105)	(0.097)	0.130
ω^*	0.0002	0.0002	0.0002	0.0002	0.0002
	(0.00006)	(0.00006)	(0.00006)	(0.00006)	(0.00006)
α^*	0.085	0.083	0.080	0.080	0.080
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
β^*	0.864	0.866	0.869	0.869	0.869
	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)
$\hat{l}$	2972.6	2983.2	2996.5	3002.9	3018.5
AIC	-5935.3	-5952.4	-5979.0	-5991.7	-6022.9
BIC	-5907.2	-5913.2	-5939.8	-5952.6	-5983.8

Robust standard errors in parentheses. ϕ_0^* denotes the constant in the conditional mean, ϕ_1^* is the AR-parameter, ϕ_{Ra}^* and $\phi_{A\_R}^*$ are the coefficients of the range and absolute return and ω^*, α^* and β^* are parameters of the GARCH part of the models.

In the literature, to the best of our knowledge, only elliptical copulas have been applied to VMEM modeling. However, it is plausible that elliptical copulas are not able to capture the highly asymmetric nature of the dependence of various variables, such as different volatility indicators.

Again, to the best of our knowledge, the number of copulas that allow for the type of asymmetry required in our application is extremely limited. The copula considered is a generalization of a certain copula that belongs to the class of Archimedean copulas. The main restriction of the approach taken in the thesis is that in the presented form it is only suitable for a bivariate application. Research where proposed methods are generalized to allow for more variables is obviously needed.

In the application considered to absolute returns and daily price range of the S&P 500 index notable advantage in the fit of the model was found after employing the asymmetric copula-function. Improvement was detected in terms of log-likelihood function, standard model selection criteria, chi-square statistics, and visual inspection of the probability integral transformed data. Most notably, a substantial improvement was discovered when the conditional distribution of the variables was examined given the realized value of the second variable. Inverse gamma distribution proposed for the daily price range in the preceding chapter of the thesis, was again found satisfactory for the range. We also provided additional evidence for the previously established result that the range has explanatory power for other volatility indicators.

It was also discovered that different copulas lead to slightly different estimates of the mean parts of the models. Obviously, even if the specified models were of the same order, different estimates lead to different forecast dynamics. We examined the effects of these changes with an application where the so-called VIX index (first differences) was modeled with a model specification where suitable multiperiod volatility forecasts from various VMEM models were considered as explanatory variables. In the application the VIX index can be treated as a market based volatility indicator as it is based on option prices whereas VMEM model based volatility indicators were considered as model based volatility indicators. The results showed that the copula used was related to the explanatory power of the forecasts such that forecasts from our asymmetric VMEM model were the best explanatory variables both in-sample and out-of-sample.

As already noted, generalizations of the proposed methods to higher dimensional systems are an obvious topic for future work. The analysis could be extended to include variables such as realized volatility and realized range and their various alternatives. In our opinion, the relations of different volatility indicators and, in particular, model and market-based volatility indicators could also be more thoroughly investigated. The applications presented in this thesis as well as in the literature suggest that model based volatility forecasts carry additional information

for market based volatility indicators. Thus, taking this information into account could enable even more accurate volatility forecasts.

References

Aas, K. 2004: Modelling the dependence structure of financial assets: A survey of four copulas, SAMBA/22/04.

Aas, K., C. Czado, A. Frigessi & H. Bakken 2009: Pair-copula constructions of multiple dependence, *Insurance: Mathematics & Economics*, 44, 182-198.

Ahoniemi, K. 2006: Modeling and forecasting implied volatility - An econometric analysis of the VIX index, HECER Discussion Paper No. 129.

Ahoniemi, K. & M. Lanne 2009: Joint modeling of call and put implied volatility, *International Journal of Forecasting*, 25, 239-258.

Bollerslev, T. 1986: Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 31, 301-327.

Bouyé, E., V. Durrleman, A. Nikeghbali, G. Riboulet & T. Roncalli 2000: Copulas for finance: A reading guide and some applications, Technical report, Groupe de Recherche Opérationelle, Crédit Lyonnais, Paris.

Chou, R. Y. 2005: Forecasting financial volatilities with extreme values: The conditional autoregressive range (CARR) model, *Journal of Money, Credit and Banking*, 37, 561-582.

Cipollini, F., R. F. Engle & G. M. Gallo 2006: Vector multiplicative error models: Representation and inference, NBER Working Paper Series, 12690.

Cipollini, F., R. F. Engle & G. M. Gallo 2007: A model for multivariate non-negative valued processes in financial econometrics, Working Papers (2007/16), DSUNIFI, Dipartimento di Statistica "Giuseppe Parenti".

Diebold, F. X., J. Hahn & A. S. Tay 1999: Multivariate density forecast evaluation and calibration in financial risk management: High-frequency returns on foreign exchange, *The Review of Economics and Statistics*, 81, 661-673.

Embrechts, P., F. Lindskog & A. McNeil 2001: Modelling dependence with copulas, Technical report, Department of Mathematics, ETHZ, CH-8092 Zurich Switzerland.

Engle, R. F. 2002: New frontiers for ARCH models, *Journal of Applied Econometrics*, 17, 425-446.

Engle, R. F. & G. M. Gallo 2006: A multiple indicators model for volatility using intra-daily data, *Journal of Econometrics*, 131, 3-27.

Engle, R. F. & J. E. Russel 1998: Autoregressive conditional duration: A new model for irregularly spaced transaction data, *Econometrica*, 66, 1127-1162.

Fleming, J., B. Ostdiek & R. E. Whaley 1995: Predicting stock market volatility: A new measure, *The Journal of Futures Markets*, 15, 265-302.

Hurd, M., M. Salmon & C. Schleicher 2007: Using copulas to construct bivariate foreign exchange distributions with an application to the sterling exchange rate index, Bank of England Working Paper no. 334.

Mari, D. & V. Monbet 2004: Construction de copules asymétriques avec dépendance extrême. Application à des données d'états de mer, Mimeo, Laboratoire Sabres.

Nelsen, R. B. 1999: An Introduction to Copulas. New York: Springer.

Sklar, A. 1996: Random variables, distribution functions, and copulas - A personal look backward and forward, in L. Rüschendorff, B. Schweizer & M. Taylor, ed., Distributions with Fixed Marginals and Related Topics, Institute of Mathematical Statistics, Hayward, CA, 1-14.

Venter, G. G. 2001: Tails of copulas, in Proceedings ASTIN Washington, USA, 68-113.

Yue, S., T. B. M. J. Ouarda & B. Bobée 2001: A review of bivariate gamma distributions for hydrological application, *Journal of Hydrology*, 246, 1-18.