Modeling the Dynamics of Stock Prices Using Realized Variation Measures

Inaugural-Dissertation
zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften
durch die
Rechts- und Staatswissenschaftliche Fakultät
der Rheinischen Friedrich-Wilhelms-Universität
Bonn

vorgelegt von
Uta Pigorsch
aus Abstatt.

Bonn 2007

Dekan:	Prof. Dr. Gerhard Wagner
Erstreferent:	Prof. Dr. Jörg Breitung
Zweitreferent:	Prof. Dr. Erik Theissen

Tag der mündlichen Prüfung: 30. August 2007

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn

`http://hss.ulb.uni-bonn.de/diss_online`

elektronisch publiziert.

Contents

List of Tables

List of Figures

1 Introduction

Modeling the dynamics of stock prices is a major theme in financial econometrics, as it has important implications for investment and risk management decisions, portfolio choices and derivative pricing. As a consequence, a plethora of different models has been developed aiming at capturing the stylized facts of stock returns. Given the clustering in the volatility of financial returns, modeling the volatility dynamics is one of the most important issues in this literature. Modeling volatility, however, is complicated by its unobservability, leading to the development of the well–known class of generalized autoregressive conditionally heteroscedastic (GARCH) models and the stochastic volatility models, in which the volatility is either a deterministic function of past returns or follows a latent stochastic process, respectively.

Recently, the availability of high–frequency financial data has opened new research directions in this field. In particular, building on the theory of quadratic variation, the high–frequency returns can be used to construct ex–post lower frequency non–parametric and consistent measures of the variation of the price process, such as the realized variance, which is a consistent measure of the quadratic variation, or the realized Bipower variation measuring only the continuous sample path variation (see e.g. Andersen and Bollerslev (1998), Andersen et al. (2001b) and Barndorff-Nielsen and Shephard (2002b, 2004b, 2005)). As such the measures provide new and useful information on the dynamics of stock prices, for example it allows to assess the existence and relevance of price jumps. Moreover, based on these measures the volatility is now treated as an observed rather than latent variable to which standard time series procedures can be applied. Examples of such realized volatility models include Andersen et al. (2003), Martens et al. (2004), Martens and Zein (2004), Pong et al. (2004), and Thomakos and Wang (2003), among others, who have advocated the use of autoregressive fractionally integrated moving average (ARFIMA) models, along with approximate long–memory component type structures in Andersen et al. (2007) and Corsi (2004). In addition, the realized volatility also allows to assess the predictive performance of the different volatility models, showing that the realized volatility models generally outperform the conventional stochastic volatility or GARCH–type models.

In this thesis, we therefore make use of the realized variation measures for modeling the individual as well as the cross–sectional dynamics of stock returns. We extend the existing literature in several respects. After a brief review of the realized variation measures and their empirical properties (see Chapter 2), we first show in Chapter 3 that the residuals of the most commonly used realized volatility models exhibit volatility clustering and non–Gaussianity. Given this observation, the usually imposed assumption of identically and independently Gaussian distributed

innovations seems to be inadequate leading potentially to inefficiencies in the estimation of such realized volatility models and to distortions in their predictive ability, in turn impairing risk management. We therefore propose two model extensions that explicitly account for the time–variation in the *volatility of the realized volatility* as well as for the non–Gaussianity, and assess their relevance for modeling and forecasting volatility.

Following the existing literature, the above approach models realized volatility, i.e. formulates a model for the total price variation, and thus makes no distinction between the different sources of volatility, that is whether the variation comes from the continuous–sample path evolvement of the price process or from price jumps. In fact, answering the question of whether the price process exhibits jumps is important for an adequate modeling and risk management of financial assets, as well as for derivative pricing. Based on the high–frequency data such distinction becomes feasible. In particular, the realized Bipower variation measure allows for a direct non–parametric decomposition of the total price variation into its two separate components, i.e. the continuous sample path variation and the variation coming from jumps. Utilizing these ideas, Andersen et al. (2007) and Huang and Tauchen (2005) both report empirical evidence in support of a significant contribution of jumps to total price variation. The simplicity of the existing reduced–form realized volatility models therefore comes at the cost of disregarding information about the relevance of jumps and the distinctly different dynamic and distributional features of the two volatility components. At least to our knowledge, a correspondingly more structured approach has not yet been considered in the literature. In Chapter 4 we take this avenue and develop an empirically highly accurate simultaneous equation model for the returns, the realized continuous sample path and the jump variation measures. The joint modeling allows us to assess the structural inter–dependencies among the shocks to returns and the two different volatility components. Moreover, we can investigate whether the often observed asymmetry between returns and volatility, oftentimes referred to as the "leverage effect", works through the continuous volatility component and/or the jumps.

So far, the existing realized volatility literature has primarily focused on measuring and modeling the volatility of a single financial asset. However, the joint modeling of the cross–sectional dynamics of multiple assets is also of major importance for risk management applications and portfolio allocation. To this end, the multivariate realized covariance measure as first introduced by Barndorff-Nielsen and Shephard (2004a) does not only provide new and useful information, but also allows to consider larger portfolios as is generally the case with daily data. This is due to the fact that, as long as the number of assets does not exceed the sampling frequency, at which the realized covariance measure is constructed, the realized covariance measure is positive definite. In Chapter 5 we therefore exploit the realized covariance measure and its information for modeling the joint dynamics of stock prices. Our approach is novel as we no longer assume that the true covariance is observable — as is the case in the existing discrete–time realized (co)variance models — and as we do not specify a purely latent covariance process. Instead we propose

a multivariate discrete–time generalized hyperbolic stochastic volatility model, in which the mean of the unobserved "true" covariance depends on the lagged realized covariances. In doing so we acknowledge the fact that in practice, once market microstructure effects have been accounted for, the realized covariance is certainly an unbiased but importantly a noisy estimator of the quadratic covariation. Based on the univariate model specification we also show that the most popular GARCH models can be represented as Gaussian stochastic volatility models where the mean of the variance is modeled rather than the variance itself — as is usually claimed in the GARCH literature.

The remainder of this thesis is organized as follows. The next chapter provides a short review of the relevant theory and a discussion on the construction of the different realized (co)variation measures. The empirical properties of the univariate measures are illustrated for S&P500 index futures, whereas the realized covariance is constructed for three component stocks of the S&P500 index. The datasets are used in the empirical application of the univariate and multivariate realized volatility models, respectively. In Chapter 3, which is based on the paper by Corsi et al. (2007), we propose a realized volatility model that accounts for non–Gaussianity and time–variation in the volatility of realized volatility. Moreover, we investigate the importance of these features for modeling and forecasting realized volatility. Chapter 4 is given by Bollerslev et al. (2007) and develops the stochastic volatility model for the joint dynamics and inter–dependencies of returns, the jump and the continuous–time component of total price variation. Chapter 5 introduces a multivariate Gaussian stochastic volatility model, in which the mean of the latent covariance depends on the realized covariance measure. Chapter 6 finally concludes.

2 Realized Variation Measures

The availability of high–frequency financial data revealed new information on the price process that can be used for developing more accurate models for the dynamics of asset returns. On the tick–by–tick level, however, modeling is complicated by specific intradaily patterns, such as the U–shape in the intraday volatility, lunch–break or market microstructure effects. Using instead the lower frequency, e.g. daily, realized variation measures alleviates these effects while retaining the most important information contained in the high–frequency data. In particular, building on the theory of quadratic variation the sum over products of intraday returns provides ex–post non–parametric measures of the variation and covariation of the price process, which is of major importance for many financial applications. Moreover, utilizing that different types of products result in measures being more or less robust to the occurrence of jumps, the individual price process can be non–parametrically decomposed into the variation coming from its continuous sample path evolvement and into the variation stemming from the jumps. As such, the measures do not only provide new information on the dynamics of the two volatility components, but also allow to assess their interrelationships among each other and with the returns.

In practice, the realized (co)variation measures are unbiased but noisy estimators of the variation of the price process. Hence, the uncertainty associated with these estimators should not be neglected. To gain an impression of its empirical relevance, we also consider measures of the variance of the realized variance estimator.

In this chapter we review the theory and construction of the relevant univariate realized variation measures along with the realized covariance. We also provide a discussion of the data sets used in this thesis, and establish a detailed analysis of the empirical properties of the realized (co)variation series.

2.1 Theoretical Framework

This section provides a brief review of the relevant theory underlying the different variation measures employed in this thesis. A more thorough theoretical treatment can be found inter alia in Andersen et al. (2001b), Barndorff-Nielsen and Shephard (2002a), Barndorff-Nielsen et al. (2006b) and Protter (2004). We begin with a discussion of the theory of quadratic variation and its estimation using realized variance. The variance of this estimator is considered involving the notions of integrated quarticity and alternative quarticity measures, as introduced by Andersen et al. (2005, 2007) and Barndorff-Nielsen and Shephard (2003, 2004b, 2005,

2006). This enables us to compute an approximation of the volatility of the realized volatility estimator and of the volatility of the logarithmic realized variance estimator. Both measures can be used to obtain information on the time evolvement of the uncertainty associated with the realized variance estimator (see Section 2.2). Decomposing the price variation into its continuous volatility component and the variation coming from the jumps, Section 2.1.2 proceeds with a discussion of measures of both components, i.e. the Bipower variation as well as the jump measure. The realized covariance measure discussed in Section 2.1.3 provides information on the co–movements among multiple assets.

2.1.1 Realized Variance and its Volatility

In the remainder of this thesis we denote the logarithmic price of a financial asset by p_t. Note that in the presence of jumps the asymptotic distribution of the realized variance estimator has not been established in the literature yet.[1] We therefore begin our discussion by assuming that p_t follows a pure diffusion process, i.e. we start with the simplest specification. This allows us to consider in more detail the variance of the realized variance estimator (under this assumption). Note that in the next section we extend the price process by the additional inclusion of jumps.

So, let us assume that the logarithmic price is given by the following diffusion process

$$p_t = p_0 + \int_0^t \mu(s)ds + \int_0^t \sigma(s)dW(s), \tag{2.1}$$

where the mean process $\mu(t)$ is continuous and of finite variation, $\sigma(t) > 0$ denotes the càdlàg instantaneous volatility and $W(t)$ is a standard Brownian motion. Moreover, $\sigma(t)$ may follow a stochastic process as is commonly assumed in financial modeling. The theory of quadratic variation then permits the derivation of non–parametric variation measures. In particular, the quadratic variation process generally defined as

$$[p]_t = \text{plim} \sum_{j=0}^{n-1} (p_{\tau_{j+1}} - p_{\tau_j})^2, \tag{2.2}$$

where $\tau_0 = 0 \leq \tau_1 \leq \ldots \leq \tau_n = t$ denotes a sequence of partitions with $\sup_j \{\tau_{j+1} - \tau_j\} \to 0$ for $n \to \infty$, is given for the price process (2.1) as

$$[p]_t = \int_0^t \sigma^2(s)ds, \tag{2.3}$$

[1] This is partly due to the fact that the inclusion of jumps tremendously complicates such derivation. Moreover, the literature has primarily focused on deriving a test statistic and its asymptotic distribution under the null of no jumps (see e.g. Andersen et al. (2007), Barndorff-Nielsen and Shephard (2006), and Huang and Tauchen (2005)).

that is, the *integrated variance*.

For the construction of the measures, let us consider the following price increments. In particular, most of our analysis will be focused on daily returns and volatilities. Hence, for notational simplicity we normalize the daily time interval to unity, denoting the corresponding daily returns by

$$r_t = p_t - p_{t-1}, \quad t = 1, \dots. \tag{2.4}$$

To formally define our empirical volatility measures, denote the day t, j–th within-day return by

$$r_{t,j} = p_{t-1+\frac{j}{M}} - p_{t-1+\frac{(j-1)}{M}}, \quad j = 1, \dots, M, \tag{2.5}$$

where M refers to the number of returns per day. The sum of the corresponding squared intradaily returns

$$RV_t = \sum_{j=1}^{M} r_{t,j}^2 \tag{2.6}$$

then affords a natural estimator of the *realized quadratic variation*. Following the recent literature we will interchangeably refer to this quantity as the *realized variance* or the *realized volatility*. The idea of measuring the ex–post variation of asset prices by summing over more frequently sampled squared returns dates back at least to Merton (1980), and was also applied by French et al. (1987), Hsieh (1991) and Poterba and Summers (1986), and more recently by Taylor and Xu (1997), inter alia. Meanwhile, the notion of realized variation was first formally related to the theory of quadratic variation within the context of finance and time–varying volatility modeling by Andersen and Bollerslev (1998), Andersen et al. (2001b), Barndorff-Nielsen and Shephard (2002b) and Comte and Renault (1998).

In particular, it follows from the theory of quadratic variation that the realized variance will generally converge uniformly in probability to the quadratic variation as the sampling frequency, M, of the underlying returns approaches infinity

$$RV_t \to \int_{t-1}^{t} \sigma^2(s) ds. \tag{2.7}$$

In other words, the realized variance provides an ex–post measure of the integrated variance.

Given a consistent estimator for the integrated variance of the price process (2.1), it is also interesting to assess the precision of this estimator. In fact, the asymptotic distribution of the realized variance has been derived in Barndorff-Nielsen and Shephard (2002a,b, 2003, 2004a, 2005) and is given by

$$\frac{\sqrt{M}\left(RV_t - \int_{t-1}^{t} \sigma^2(s) ds\right)}{\sqrt{2\int_{t-1}^{t} \sigma^4(s) ds}} \xrightarrow{d} N(0,1),$$

where $\int_{t-1}^{t} \sigma^4(s)ds$ denotes *integrated quarticity*.

Unfortunately, however, the computation of the asymptotic distribution is infeasible, given that the integrated quarticity is unknown. Based on the theory of power variation, Barndorff-Nielsen and Shephard (2002a, 2004b, 2006) suggest different estimators of this quantity. E.g. the *realized fourth-power variation* or *realized quarticity*, defined as

$$RQ_t = \frac{M}{3}\sum_{j=1}^{M} r_{t,j}^4 \rightarrow \int_{t-1}^{t} \sigma^4(s)ds, \tag{2.8}$$

is a consistent estimator of the integrated quarticity. Alternative estimators are the *realized quad-power quarticity*,

$$RQQ_t = M\frac{\pi^2}{4}\sum_{j=4}^{M} |r_{t,j}||r_{t,j-1}||r_{t,j-2}||r_{t,j-3}| \rightarrow \int_{t-1}^{t} \sigma^4(s)ds. \tag{2.9}$$

and the *realized tri-power quarticity*,

$$RTQ_t = M\frac{\Gamma\left(\frac{1}{2}\right)^3}{4\Gamma\left(\frac{7}{6}\right)^3}\sum_{j=3}^{M} |r_{t,j}|^{\frac{4}{3}}|r_{t,j-1}|^{\frac{4}{3}}|r_{t,j-2}|^{\frac{4}{3}} \rightarrow \int_{t-1}^{t} \sigma^4(s)ds, \tag{2.10}$$

proposed in Andersen et al. (2007), where Γ denotes the Gamma function.

Based on these different quarticity measures, the asymptotic distribution of realized variance can be approximated by

$$\frac{RV_t - \int_{t-1}^{t} \sigma^2(s)ds}{\sqrt{\frac{2}{M}Q_t^*}} \xrightarrow{d} N(0,1), \quad Q_t^* \in \{RQ_t, RQQ_t, RTQ_t\}, \tag{2.11}$$

where $\sqrt{\frac{2}{M}Q_t^*}$ provides an approximation of the standard deviation of the realized variance error.

As we are interested in the volatility of the realized–volatility error, we apply the delta method and obtain as an approximation of the standard deviation of the realized volatility estimator $\sqrt{\frac{Q_t^*}{2MRV_t}}$ with $Q_t^* \in \{RQ_t, RQQ_t, RTQ_t\}$. Hence, using the different measures of integrated quarticity, we can compute three alternative approximations of the (daily) volatility of the realized volatility estimator, namely,

$$\sqrt{\frac{RQ_t}{2MRV_t}} = \sqrt{\frac{\sum_{j=1}^{M} r_{t,j}^4}{6\sum_{j=1}^{M} r_{t,j}^2}} \tag{2.12}$$

$$\sqrt{\frac{RQQ_t}{2MRV_t}} = \sqrt{\frac{\pi^2\sum_{j=4}^{M} |r_{t,j}||r_{t,j-1}||r_{t,j-2}||r_{t,j-3}|}{8\sum_{j=1}^{M} r_{t,j}^2}} \tag{2.13}$$

$$\sqrt{\frac{RTQ_t}{2MRV_t}} = \sqrt{\frac{\Gamma\left(\frac{1}{2}\right)^3\sum_{j=3}^{M} |r_{t,j}|^{\frac{4}{3}}|r_{t,j-1}|^{\frac{4}{3}}|r_{t,j-2}|^{\frac{4}{3}}}{8\Gamma\left(\frac{7}{6}\right)^3\sum_{j=1}^{M} r_{t,j}^2}}. \tag{2.14}$$

As the logarithmic transformation of the realized variance is well–known to provide better finite sample properties (see e.g. Barndorff-Nielsen and Shephard (2005)), we are also interested in the volatility of the logarithmic realized–variance error. In particular, Barndorff-Nielsen and Shephard (2001) derive the approximate asymptotic distribution of the logarithmic transform of realized variance

$$\frac{\log RV_t - \log \int_{t-1}^{t} \sigma^2(s)ds}{\sqrt{\frac{2Q_t^*}{M(RV_t)^2}}} \xrightarrow{d} N(0,1), \quad Q_t^* \in \{RQ_t, RQQ_t, RTQ_t\}. \tag{2.15}$$

Based on the realized quarticity measure, Barndorff-Nielsen and Shephard (2006) show in a simulation study that the implied standard errors for the logarithmic transformation indeed tend to be smaller than those of the realized variance statistic given. For both measures, the integrated variance is estimated by the realized quarticity measure. The different measures of the (daily) volatility of the logarithmic realized–variance error are given by

$$\sqrt{\frac{2RQ_t}{M(RV_t)^2}} = \sqrt{\frac{2}{3}\frac{\sum_{j=1}^{M} r_{t,j}^4}{(\sum_{j=1}^{M} r_{t,j}^2)^2}} \tag{2.16}$$

$$\sqrt{\frac{2RQQ_t}{M(RV_t)^2}} = \sqrt{\frac{\pi^2}{2}\frac{\sum_{j=4}^{M} |r_{t,j}||r_{t,j-1}||r_{t,j-2}||r_{t,j-3}|}{(\sum_{j=1}^{M} r_{t,j}^2)^2}} \tag{2.17}$$

$$\sqrt{\frac{2RTQ_t}{M(RV_t)^2}} = \sqrt{\frac{\Gamma(\frac{1}{2})^3}{2\Gamma(\frac{7}{6})^3}\frac{\sum_{j=3}^{M} |r_{t,j}|^{\frac{4}{3}}|r_{t,j-1}|^{\frac{4}{3}}|r_{t,j-2}|^{\frac{4}{3}}}{(\sum_{j=1}^{M} r_{t,j}^2)^2}}. \tag{2.18}$$

The empirical properties of these "*volatility of realized volatility*" measures as well as of the realized variance are illustrated in Section 2.2.2.

2.1.2 Bipower Variation and Jumps

In this section we extend the price process given in equation (2.1) by additionally including a jump process. This case is especially interesting. In particular, if the price process exhibits jumps, the theory of quadratic variation allows us to disentangle empirically the variation coming from the diffusion component and that of the jumps by constructing non–parametric measures of these two components. Hence, we assume in the following that p_t follows the continuous–time semimartingale jump diffusion process:

$$p_t = \int_0^t \mu(s)ds + \int_0^t \sigma(s)dW(s) + \sum_{j=1}^{N(t)} \kappa(s_j), \tag{2.19}$$

where $N(t)$ denotes a process that counts the number of jumps occurring with possibly time–varying intensity and jump size $\kappa(s_j)$. For this process the quadratic

variation (as defined in (2.2)) turns out to be

$$[p]_t = \int_0^t \sigma^2(s)ds + \sum_{j=1}^{N(t)} \kappa^2(s_j), \tag{2.20}$$

that is, the *integrated variance* and the sum of the *squared jumps*, which is the variation coming from the jumps.

Thus, in this case, the realized variance as defined in equation (2.6) consistently estimates the total price variation

$$RV_t \to \int_{t-1}^t \sigma^2(s)ds + \sum_{j=N(t-1)+1}^{N(t)} \kappa^2(s_j), \tag{2.21}$$

i.e. including the discontinuous jump part.

In order to distinguish the continuous variation from the jump component, Barndorff-Nielsen and Shephard (2004b) first proposed the so-called *Bipower variation* measure, defined by

$$BV_t = \frac{\pi}{2} \sum_{j=2}^{M} |r_{t,j}||r_{t,j-1}|. \tag{2.22}$$

Importantly, for increasingly finely sampled returns the Bipower variation measure becomes immune to jumps and consistently (for increasing values of M) estimates the integrated variance

$$BV_t \to \int_{t-1}^t \sigma^2(s)ds. \tag{2.23}$$

Consequently, the difference between the realized variance and the Bipower variation affords a simple non–parametric estimator of the contribution to total price variation coming from the jump component, i.e. $JL_t = RV_t - BV_t$. Motivated by the availability of these measures, we assess their empirical properties in the next section. Moreover, Chapter 4 focuses on modeling their dynamics and interrelationships with daily returns. Meanwhile, given the aforementioned better finite sample performance of the logarithmic transform of the series, we primarily base our analysis on their logarithmic versions and consider the following jump measure

$$J_t = \log RV_t - \log BV_t \tag{2.24}$$

having support over the whole real line. Note that in practice the measurement errors in the realized variance and Bipower variation measures translate into the jump series. Thus small values of the jump measure can partly be attributed to measurement errors, and, thus, are not necessarily associated with "true" jumps. Alternatively, building on the asymptotic (for increasing M) distribution theory

in Barndorff-Nielsen and Shephard (2004b), it would also be possible to truncate the J_t process, and only associate the values beyond a certain threshold with the jump component. This is the approach adopted in Andersen et al. (2007), who rely on a large critical value for identifying only the most significant jumps entering a reduced form univariate forecasting model for RV_t. In contrast, our approach avoids the arbitrary choice of any pre–specified significance level affecting the selection of "significant" jumps.

2.1.3 The Realized Covariance Measure

We now consider the joint process of multiple asset returns. Let $\boldsymbol{p}_t$ denote the $d \times 1$ vector of the logarithmic prices of d different assets. Assume for simplicity that it follows a d–dimensional pure diffusion process similarly to the univariate case given in equation (2.1), i.e.

$$\boldsymbol{p}_t = \boldsymbol{p}_0 + \int_0^t \boldsymbol{\mu}(s)ds + \int_0^t \boldsymbol{\Sigma}^{1/2}(s)d\boldsymbol{W}(s) \tag{2.25}$$

with d–dimensional drift $\boldsymbol{\mu}(t)$, $\boldsymbol{\Sigma}(t) = \boldsymbol{\Sigma}^{1/2}(t)\boldsymbol{\Sigma}^{1/2}(t)'$ denoting the instantaneous covariance, and $\boldsymbol{W}(t)$ is a d–dimensional standard Brownian motion. Moreover, denote the vector of the j–th intraday returns of day t by $\boldsymbol{r}_{t,j}$ and define the daily realized covariance measure by

$$\boldsymbol{RC}_t = \sum_{j=1}^{M} \boldsymbol{r}_{t,j}\boldsymbol{r}'_{t,j}. \tag{2.26}$$

Then, for $M \to \infty$ the realized covariance measure converges uniformly in probability to the daily increments of the quadratic covariation, which is given here by the daily integrated covariance matrix, i.e.

$$\boldsymbol{RC}_t \to \int_0^t \boldsymbol{\Sigma}^{1/2}(s)ds, \tag{2.27}$$

as shown in Barndorff-Nielsen and Shephard (2004a), who also derive further asymptotic properties of the realized covariance estimator.

2.2 Implementation and Empirical Properties

This section introduces the data sets that are used in the empirical applications of this thesis. Issues on the practical implementation of the above described measures are also discussed. Their empirical properties are illustrated in Section 2.2.2.

2.2.1 Data and Construction

Our empirical applications of the models developed in Chapters 3 and 4 are based on tick–by–tick transaction prices of S&P500 index futures recorded at the Chicago Mercantile Exchange (CME). The sample covers the period from January 1, 1985 to December 31, 2004, a period of 5,040 trading days, and consists of 13,241,032 tick–by–tick observations. Note that we disregard the overnight trading of contracts at GLOBEX, the CME overnight trading platform, as it just started in 1994. The multivariate model of Chapter 5 instead is applied to three S&P500 component stocks. In particular, we use the tick–by–tick transaction prices of the stocks of Intel, Microsoft and Pfizer. The data is taken from the Trade and Quote (TAQ) Database and ranges from January 1, 2001 to December 31, 2005. Note that we only consider valid trades taking place during the official trading time, i.e. from 9.30 a.m. to 4 p.m.[2] We, thus, exclude trading overnight.

When it comes to the construction of the different realized variation measures, several problems arise in practice. Recall, that the theory discussed in the preceding section formally hinges on the notion of increasingly finer sampled high–frequency returns. In practice, however, the sampling frequency is invariably limited by the actual quotation, or transaction frequency. Moreover, the observed high–frequency prices are further "contaminated" by a host of market microstructure frictions, including price discreteness, transaction costs and bid–ask spreads. These effects lead to important biases in the realized variation measures. Thus, a bias–variance trade–off has to be solved. On one hand, the sampling frequency should be as small as possible in order to obtain precise measures, on the other hand the bias due to the prevalent market microstructure effects becomes more severe if the sampling frequency increases. In response to this, a number of authors, including Andersen et al. (2001a,b, 2007), have advocated the use of coarser sampling frequencies, such as five to 30 minutes, as a simple way to alleviate these contaminating effects, while maintaining most of the relevant information in the high-frequency data. This is also the approach adopted here. Note that several recent studies have proposed alternative procedures to more effectively make use of all the tick–by–tick data including the notion of an optimal sampling frequency, in the sense of minimizing the MSE of the resulting realized volatility measure as suggested by Aït-Sahalia et al. (2005) and Bandi and Russell (2005a), or business type sampling schemes dictated by the activity of the market, as in, e.g., Oomen (2005). Other studies adopt techniques that are usually applied in the estimation of the variance of a stationary time series in the presence of autocorrelation. The reason behind this is that the market microstructure noise leads to autocorrelation in the intraday returns. Examples of such approaches include the use of various pre-filtering or kernel type procedures, e.g., Andersen et al. (2001a), Bollen and Inder (2002), Corsi et al. (2001), Hansen and Lunde (2006), and Zhou (1996). Zhang et al. (2005) were the first to advocate sub–sampling schemes designed to adjust for the bias in the

[2]The valid trades are already classified by the TAQ database. In particular, we do not consider trades that are indicated by a "exclude" and "error" flag.

simple realized volatility estimator for increasing values of M. The recent paper by Barndorff-Nielsen et al. (2006a) provides a unified theoretical framework for analyzing most of these estimators within a kernel based representation, along with a discussion of optimal kernel and bandwidth choices. Meanwhile, to the best of our knowledge none of these ideas have yet been formally extended to allow for similar measurements of the integrated variance or integrated quarticity in form of robust to market microstructure noise modified realized Bipower variation or realized quarticity measures, respectively. Hence, in the empirical results reported in Chapters 3 and 4, we simply rely on the same coarse sampling interval in the construction of all measures.

Similar problems arise in the construction of the realized covariation measure, which is additionally complicated by the non–synchronous trading of the different assets as well as by the possibility of cross–correlated market microstructure noise. In empirical applications the data is therefore oftentimes synchronized, e.g. using the last–tick interpolation, which, however, introduces an additional so–called synchronization bias. As a consequence, Bandi and Russell (2005b), Hayashi and Yoshida (2005) and Voev and Lunde (2007) developed different methods for correcting the realized covariances for this synchronization–bias. However, at least to our knowledge, a joint approach for correcting for the different sources of the biases inherent in the realized covariance measure is still pending in the literature. Moreover, such approach may be complicated by the potentially different impacts of market microstructure noise on the realized covariances relative to the realized variances (see also the discussion in Voev and Lunde (2007)). As such, the multivariate extension of the existing correction methods for the realized variance may not be straightforward. Based on this lack of a unifying approach for correcting jointly the full realized covariance matrix, we follow Barndorff-Nielsen and Shephard (2004a) and compute the realized covariance matrix according to equation (2.26) using 30–minute returns and by applying the last–tick interpolation.

So, the precise construction of our realized variation and covariation measures is as follows. For the S&P500 index futures data, we follow Andersen and Bollerslev (1998); Andersen et al. (2001b, 2007), Maheu and McCurdy (2002), and Martens et al. (2004), among others, and use five–minute returns to construct the univariate realized variation measures. For the computation of these five–minute returns we consider only the transaction prices of the most liquid contract at the beginning of our sample period. As soon as another contract is traded more frequently we switch to that one. The corresponding intraday returns are then constructed from the transaction prices of each of these contracts and, thus, avoids the computation of returns over the roll–over period. Moreover, we use the nearest neighbor to the five–minute tag.

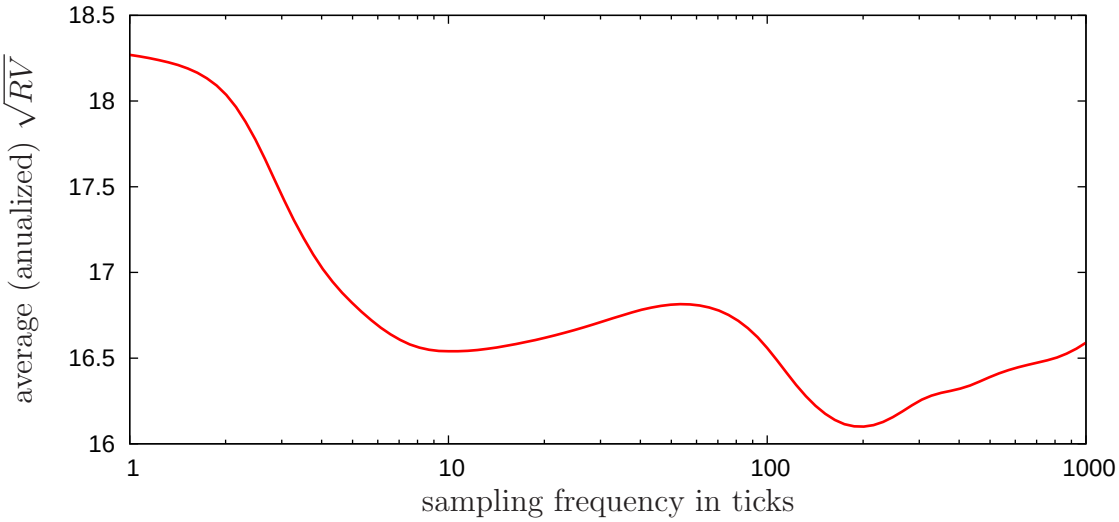

Figure 2.1: Volatility signature plot of the S&P500 index futures constructed over the full sample period. The graph shows average annualized realized volatility constructed for different frequencies measured in number of ticks. Note that there are about 7 seconds on average between trades, such that the average annualized five–minute based realized volatility corresponds to around the 43–th tick.

Importantly, the volatility signature plot given in Figure 2.1 — depicting the average annualized realized volatility over the full sample period constructed for different frequencies — indeed indicates that the market microstructure induced bias of the so constructed realized volatility is relatively small for the S&P500 index futures, and dies out very quickly. Note that with a transaction taking place on average about every seven seconds, the average annualized realized volatility based on the five–minute intervals corresponds to around the 43–th tick presented in the Figure.[3] Furthermore, note that the ratios of the sample means of the 5–minute based realized measures to the ones based on 15– and 30–minute sampling, equal 0.9936 and 0.9746 for the realized variance, and 0.9732 and 0.9660 for the Bipower variation, respectively. The impact of market microstructure effects on the five–minute realized variation measure for the S&P500 index futures over the period from 1985 to 2004 can therefore be regarded as negligible.

For our multivariate data set we compute the realized covariance measure based on 30–minute returns and by using the last–tick interpolation. As the assets considered here are highly liquid, i.e. the lowest average duration between trades is 0.33 seconds (see the last column of Table 2.2), we also expect that the market microstructure noise induced bias is negligible at this frequency.[4] Moreover, recently de Pooter et al. (2006) have found that with respect to a portfolio optimization application the optimal sampling frequency ranges between 30 to 60 minutes.

2.2.2 Empirical Properties

In the following we illustrate the empirical properties of the different realized variation and covariance measures based on the S&P500 index futures data as well as on the three S&P500 component stocks, respectively. Many of these characteristics have already been observed for other data sets and markets. Meanwhile, the detailedness of the descriptive analysis of the volatility of the realized volatility measures as well as of the two volatility components is novel to the literature, and reveals important and new insights on the dynamics and interrelationships of financial volatility. We start our analysis with a discussion of the empirical properties of the realized volatility measure and its variance. Thereafter, we present the empirical characteristics of the continuous volatility component and the jumps, and then turn to the empirical properties of the realized covariance series.

Realized Volatility and its Variance

Table 2.1 reports the descriptive statistics of the five–minute based realized volatility, $\sqrt{RV_t}$, logarithmic realized variance, $\log RV_t$, and their respective variance measures, as defined in equations (2.12) to (2.14), and in equations (2.16) to (2.18),

[3] Similarly decreasing volatility signature plots for transaction prices of liquid assets have also been shown in Hansen and Lunde (2006), for example.

[4] In fact, the ratios of the sample means of the measures based on different frequencies suggest that there exists a strong bias at higher frequencies, such as for example 15 minute returns.

respectively. Moreover, the table also presents the corresponding values of the two volatility components and daily returns, which will be discussed in the next section. The time evolvement as well as the sample autocorrelation function of the logarithmic realized volatility are presented in Figures 2.3 and 2.4 (second panels), and those of the various volatility of the realized volatility and volatility of the logarithmic realized variance measures are depicted in Figure 2.2. Table 2.1 reveals that the distribution of realized volatility is fat–tailed and slightly skewed. Taking the logarithm of realized variance leads to a strong reduction in skewness and kurtosis. Similar results for the realized volatility from other markets have previously been reported in Andersen et al. (2001a,b) among others. Given that the Gaussianity assumption is more suitable for the logarithmic transform of realized variance, Andersen et al. (2001b) model logarithmic realized volatility. However, the descriptive statistics show that skewness and kurtosis are not completely eliminated by the logarithmic transformation, as is also illustrated in Figure 2.5 showing the kernel density plots of the logarithm of realized variance (second panel).[5] Hence, the Gaussianity assumption for the innovations in the realized volatility models deserves further investigation (see Chapter 3). Gonccalves and Meddahi (2005), for example, show that other nonlinear transformations are more effective in reducing the sample skewness.

Turning to the last column of Table 2.1, the realized volatility and its logarithmic transform both exhibit highly significant autocorrelation. Furthermore, the sample autocorrelation function for the realized volatility measure depicted in the second panel of Figure 2.4 shows clearly the characteristic hyperbolic decay with autocorrelation coefficients being significant (compared to the conservative Bartlett 95% confidence bands) up to the 125th order, or roughly half–a–year. Note that this finding is no artefact of the sampling or aggregation scheme employed in the construction of the realized volatility measure, but is rather consistent with the long–memory behavior of volatility that has been extensively reported in the GARCH literature. Although the true source of long memory in volatility is not clear (see, for example, Banerjee and Urga (2005), Engle and Lee (1999) and Mikosch and Stărică (2004)), its existence has been widely recognized and captured in different volatility models, such as, for example, the FIGARCH or ARFIMA models.

The unconditional distributions of all three measures for the volatility of the realized volatility exhibit skewness and leptokurtosis, both of which are most pronounced for the realized–quarticity–based measure (2.12). This can be explained by the construction of this measure, with the fourth power yielding high values for large (absolute) intra–day returns. The autocorrelation functions presented in Figure 2.2 also show that all volatility of realized volatility series exhibit long range dependence. Moreover, the volatility of realized volatility assumes high values when realized volatility is high (cf. Barndorff-Nielsen and Shephard, 2005). Most

[5]The Kolmogorov–Smirnov test rejects the null of Gaussianity (p–value=0.0087). Our results differ from those reported in Thomakos and Wang (2003), who also perform tests on Gaussianity but use a much shorter sample period.

Table 2.1: Descriptive Statistics for the S&P500 Index Futures Data

Series	Mean	Std.Dev.	Median	Skewness	Exc.Kurt.	Ljung–Box(22)
$\sqrt{RV_t}$	0.8627	0.5935	0.7586	15.35	493.76	14,605
$\log RV_t$	-0.5139	0.8775	-0.5527	0.60	1.8070	22,023
$\sqrt{\frac{RQ_t}{2MRV_t}}$	0.0821	0.0893	0.0675	19.29	603.68	3,498
$\sqrt{\frac{RQQ_t}{2MRV_t}}$	0.0676	0.0475	0.0570	6.62	93.89	12,551
$\sqrt{\frac{RTQ_t}{2MRV_t}}$	0.0704	0.0500	0.0594	6.29	76.95	10,955
$\sqrt{\frac{2RQ_t}{M(RV_t)^2}}$	0.2034	0.0915	0.1862	12.73	298.89	56
$\sqrt{\frac{2RQQ_t}{M(RV_t)^2}}$	0.1648	0.0285	0.1599	1.71	6.75	33
$\sqrt{\frac{2RTQ_t}{M(RV_t)^2}}$	0.1720	0.0302	0.1656	2.23	9.05	85
$\sqrt{BV_t}$	0.8340	0.5359	0.7348	11.16	288.46	18,027
$\log BV_t$	-0.5817	0.8845	-0.6163	0.54	1.48	39,486
$\log\left(\frac{RV_t}{BV_t}\right)$	0.0678	0.1263	0.0538	1.78	12.27	67
r_t	0.0254	1.0946	0.0511	-2.17	96.25	125
$r_t/\sqrt{RV_t}$	0.0866	1.0027	0.0739	0.05	-0.15	25

Reported are the descriptive statistics of daily realized volatility, logarithmic realized variance, the measures of the volatility of the realized volatility estimator as defined in equations (2.12)-(2.14), the volatility of the logarithmic realized variance as defined in equations (2.16)-(2.18), as well as those of the daily Bipower variation in standard deviation form, the logarithm of the Bipower variation, of the jump measure, and of the returns and standardized returns.

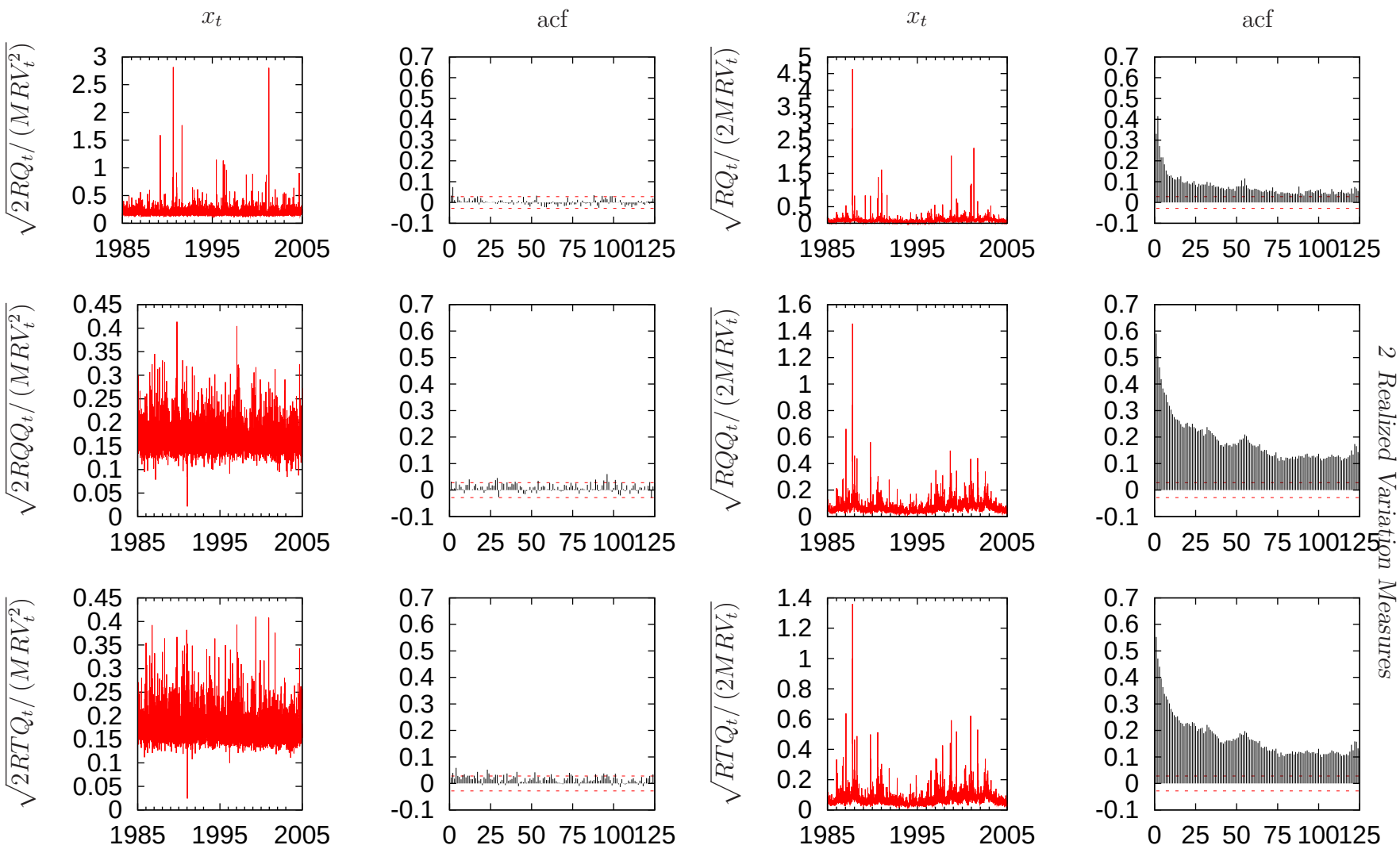

Figure 2.2: Time evolvement (first and third column panels), and sample autocorrelation functions (acf) (second and last column panels) of the three measures of the volatility of the realized volatility estimator as defined in equations (2.12)-(2.14) (first two column panels), and of the three measures of the volatility of the logarithmic realized variance estimator as defined in equations (2.16)-(2.18) (last two column panels).

importantly, all three measures exhibit time–variation and a clustering behavior, suggesting that the precision at which the realized volatility measure estimates the quadratic variation is time–varying with high precision occurring at highly precise periods and low precision following less precise periods. Similar patterns are also found for the volatility of the log–transformed realized volatility estimator. However, fluctuations over time are somewhat smaller. Note that the Ljung–Box test indicates significant autocorrelation up to at least one month for two of the three series (the critical value at the 5% significance level is 33.92).

Bipower Variation, Jumps and Returns

We now turn to the discussion of the empirical properties of the two volatility components and the returns. Their descriptive statistics are presented in the last rows of Table 2.1, while Figures 2.3 and 2.4 present their time evolvement and sample autocorrelation functions, respectively. All of the series exhibit the widely–documented volatility clustering effect. Also, comparing the (logarithmic) realized volatility with the (logarithmic) Bipower variation reveals that the variance of the former exceeds that of the continuous volatility component. Consistent with this, the jump series depicted in the last panel exhibits many, mostly positive, small values. These small observations, including the small negative values, may be attributed to measurement, or discretization, errors due to the use of finitely many returns in the construction of the underlying measures. At the same time, the series also contains a number of more extreme observations indicative of genuine large–sized jumps on those days.

These visual impressions are confirmed by the summary statistics reported in Table 2.1. In particular, the mean and variance of the realized volatility both exceed the corresponding statistics for the square–root Bipower variation. It follows also from the table that the unconditional distribution of both volatility measures are highly skewed and leptokurtic. As already noted for the realized volatility, the logarithmic transform also renders the unconditional distribution of the Bipower variation to be more or less normal. The approximate log–normality is further illustrated for both series by the kernel density plots presented in Figure 2.5. Meanwhile, the descriptive statistics and the corresponding kernel density plots for the relative jump measure, J_t, clearly indicate a positively skewed and leptokurtic distribution. The unconditional distribution of the daily returns also show the expected excess kurtosis and negative skewness. At the same time, the distribution of the returns standardized by the realized volatility is surprisingly close to Gaussian, as previously documented by Andersen et al. (2001a).[6]

Note that all of these findings strongly indicate the presence of jumps. Although we do not test for the significance of these jumps, the empirical results reported in the literature so far also strongly point towards this direction. In particular,

[6] In the absence of jumps and independence between the innovation processes driving the returns and the volatility, the standardized returns then defined by the stylized model in equation 2.1 should in fact be normally distributed.

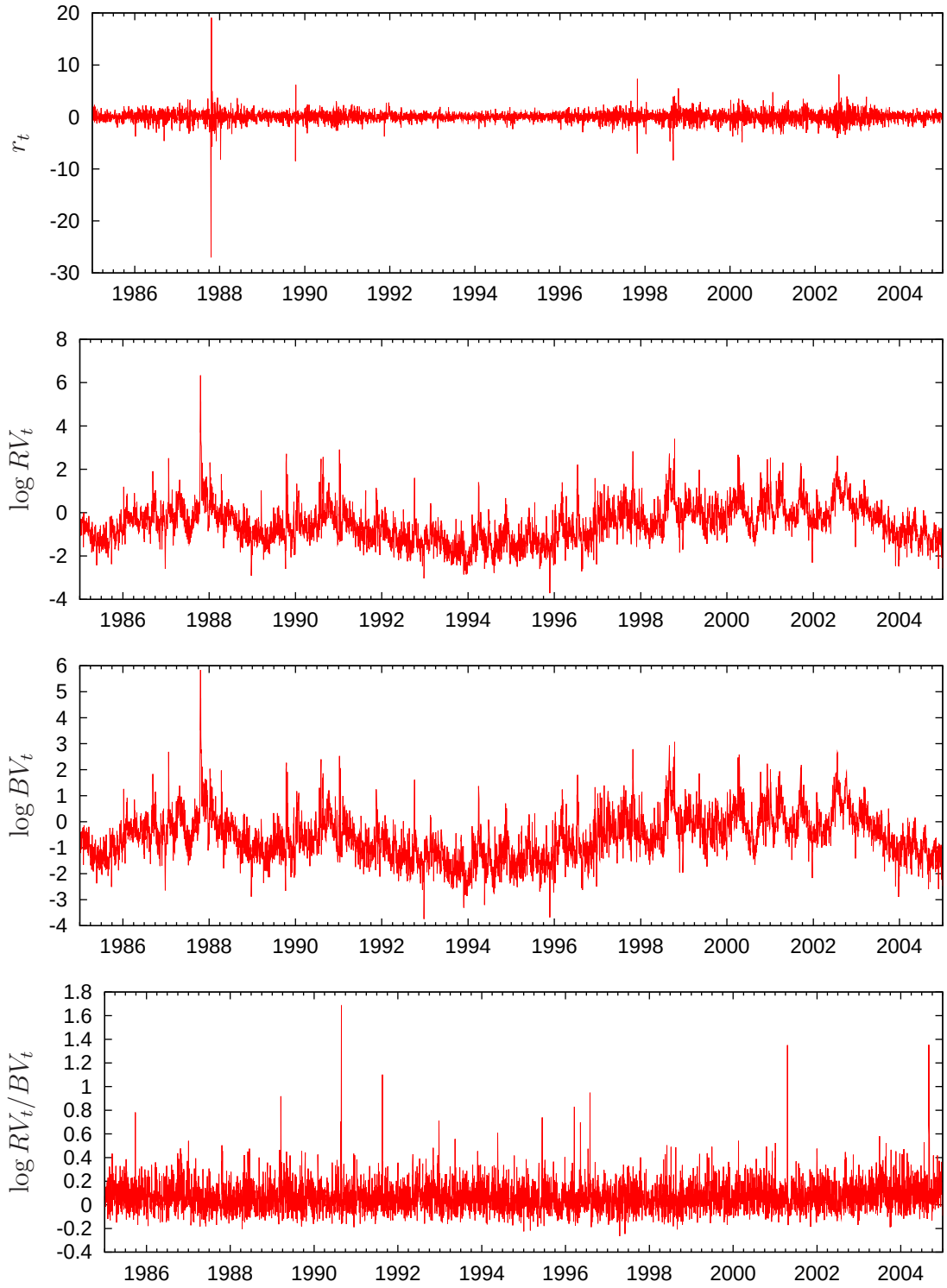

Figure 2.3: Time Series of returns, logarithmic realized variance, logarithmic Bipower variation and jumps.

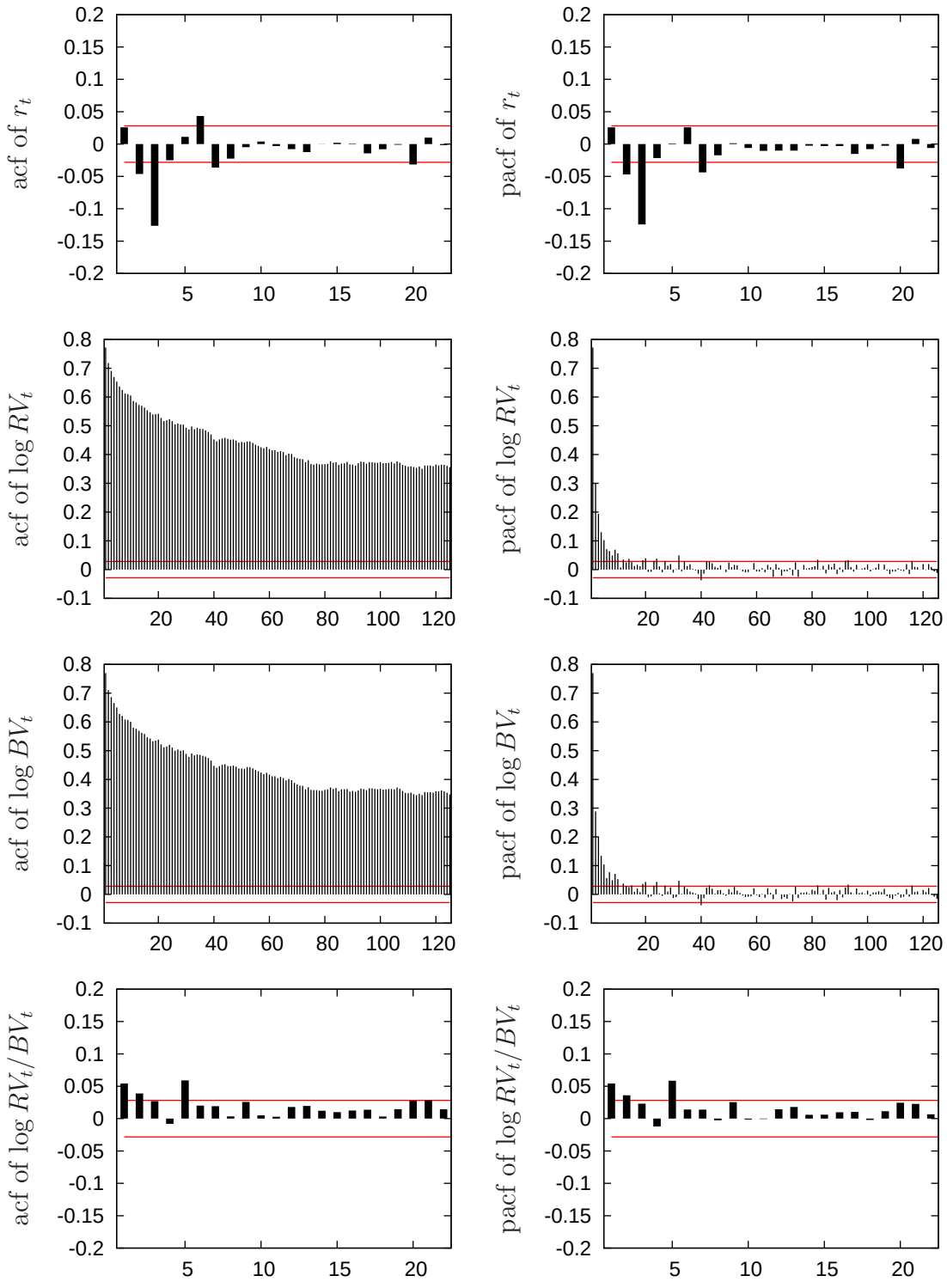

Figure 2.4: Sample autocorrelations and partial autocorrelations of returns, logarithmic realized variance, logarithmic Bipower variation and jumps. The red lines give the upper and lower ranges of the conventional Bartlett 95% confidence band.

Huang and Tauchen (2005) find for the S&P500 index that the relative contribution of jumps to total price variation amounts to roughly seven percent. As such, our realized volatility series measures the total price variation – including the jumps — rather than just the integrated variance — as we have implicitly assumed above in our discussion on the volatility of the realized variance estimator. Hence, these measures cannot anymore be strictly interpreted as the variance of the realized variance estimator. Still, as the largest part of the price variation is attributable to the continuous sample path evolvement, we argue that the time series pattern of these measures are nevertheless somewhat indicative of a volatility clustering in this error. In the subsequent modeling approaches, however, we will not rely on these measures anymore, but rather model the time–variation in the volatility of volatility — which is also observed in the residuals of the commonly employed realized volatility models — by a GARCH process.

Turning to the serial dependence of the two volatility components, we find that the Bipower variation exhibits similar highly significant autocorrelation as the realized volatility, as evidenced by the Ljung–Box test statistics for up to the 22nd order autocorrelation (representing approximately one month). This is also illustrated by the sample autocorrelation function depicted in the third panel of Figure 2.4. In contrast, the relative jump measure exhibit much less autocorrelation, with most of the dependency being attributable to the first and the fifth lag, corresponding to jumps that are one day and one week apart, respectively.

In addition to the serial correlation in the individual series, any interactions among the series will also be important in the formulation of a fully satisfactory joint model. In this regard, a number of previous studies have pointed toward a negative correlation between past return shocks and current volatility, so that "bad" news tend to be associated with a larger increase in volatility than "good" news of the same absolute magnitude.[7] A common approach for empirically visualizing this asymmetric relationship is provided by the news–impact curve originally suggested by Engle and Ng (1993). Indeed, the corresponding plots for the logarithmic realized variance and Bipower variation in Figure 2.6 both exhibit the expected slight asymmetric response to past standardized returns. The jumps, meanwhile, seem to be almost unaffected by the past return shocks, and if anything they respond *negatively* to the standardized returns. This also explains, why the asymmetric effect is more pronounced for the pure continuous volatility BV_t component in the second panel, in comparison to the total realized variation RV_t depicted in the first panel.

[7] Although this phenomenon could be explained through financial leverage, the magnitude for equity index returns is typically too large, and alternative explanations based on a time–varying volatility risk–premium have been pursued by Bekaert and Wu (2000), Campbell and Hentschel (1992), Tauchen (2005), among others. However, the causal directions of the leverage and volatility feedback effects are fundamentally different, and the recent high–frequency data analysis in Bollerslev et al. (2006b) point toward a "leverage" type causality. We will return to this issue in Section 4.2.2.

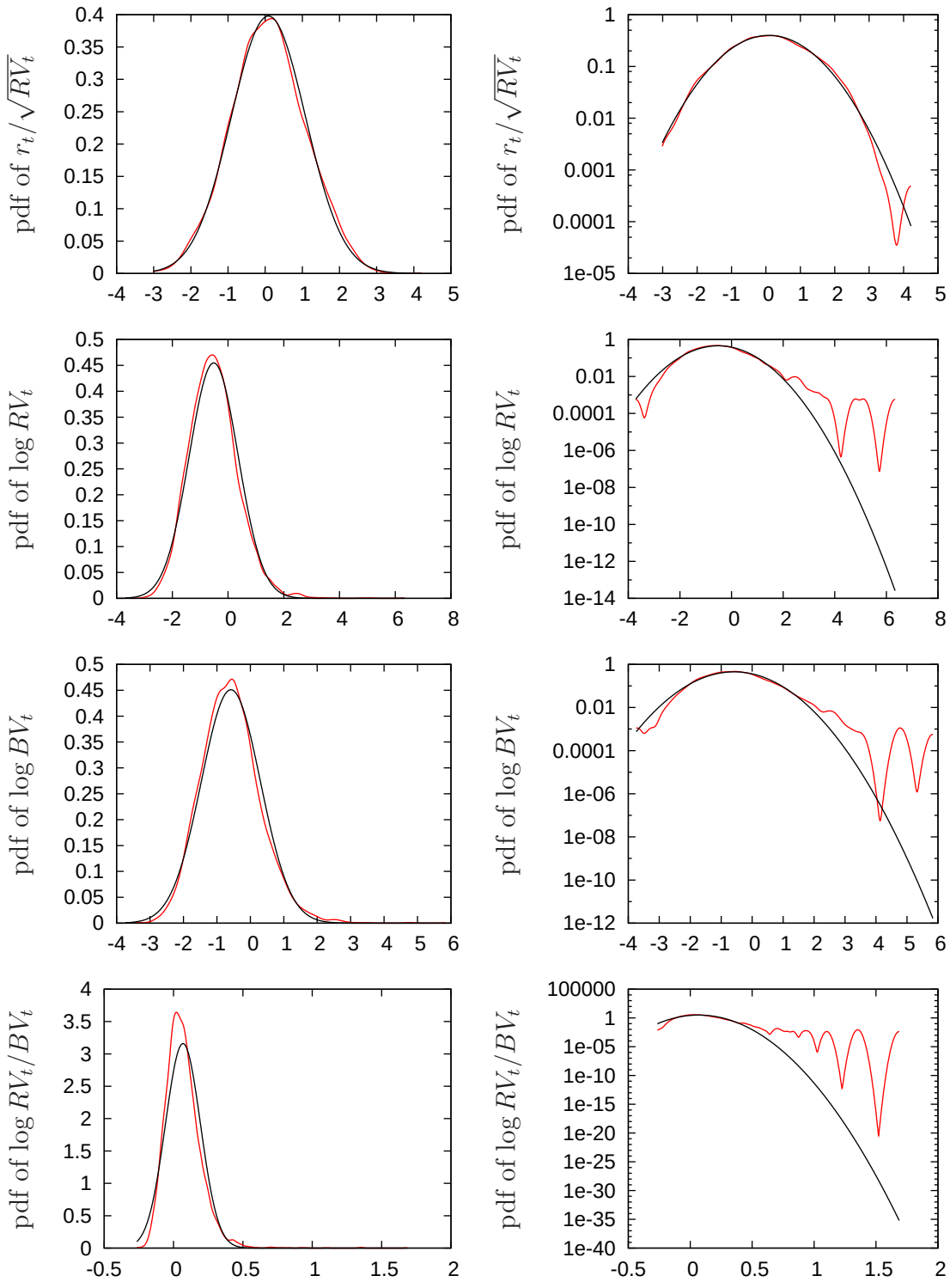

Figure 2.5: Unconditional distributions of standardized returns, logarithmic realized variance, logarithmic Bipower variation and jumps. The left panel of the figure shows the kernel density estimates of the series (red line) and the normal density (black line) for reference purposes. The right panel shows the same in log scale.

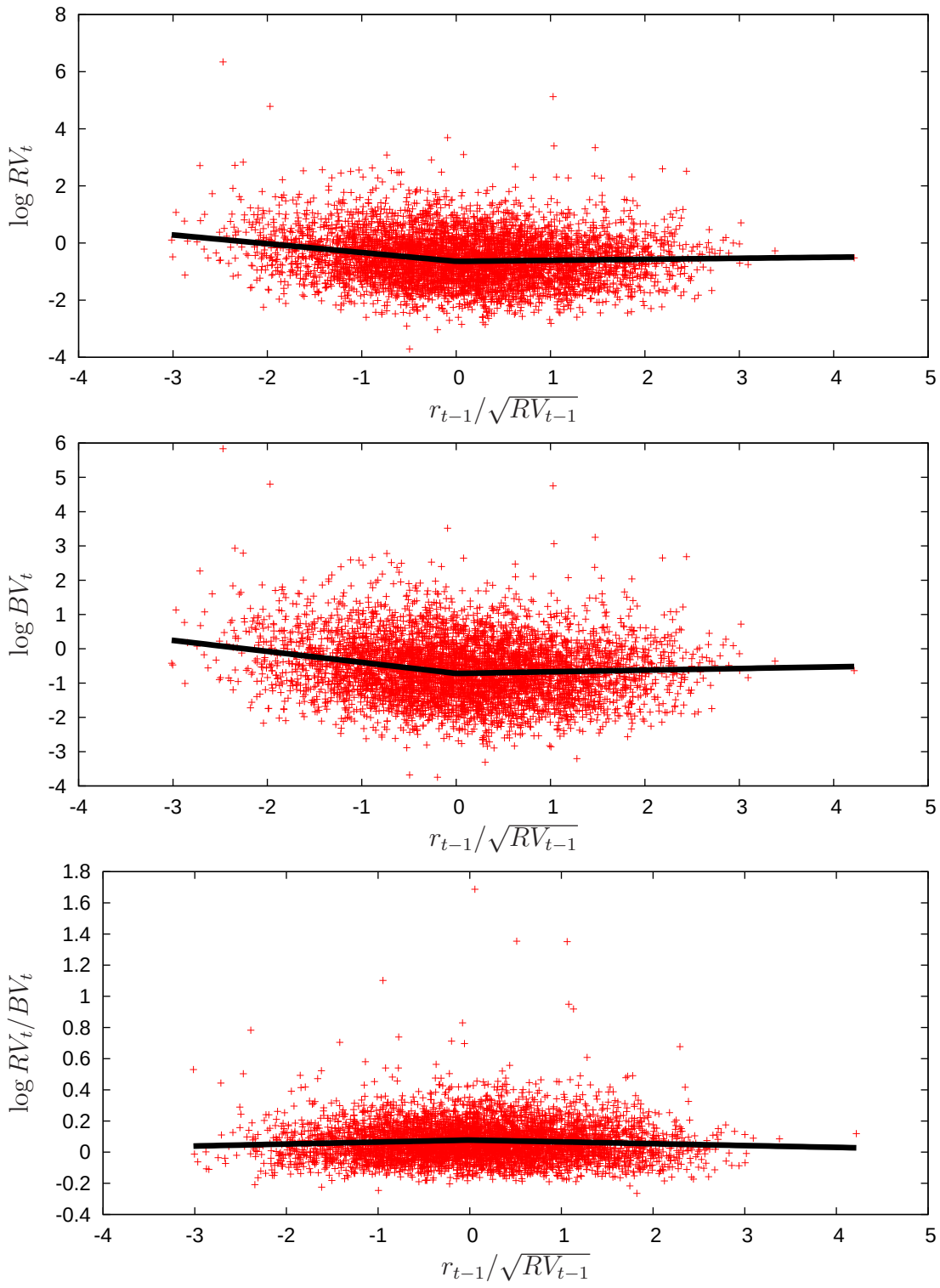

Figure 2.6: News–impact curves for logarithmic realized variance, logarithmic Bipower variation and jumps. The figure shows the scatter points between the respective variable and lagged standardized returns. The black lines are the news–impact curves, i.e. the linear regression lines for negative and positive values of standardized returns.

Table 2.2: Descriptive Statistics of the INTC, MSFT and PFE Returns

Series	Mean	Std.Dev.	Median	Skew.	Exc.Kurt.	LB(22)	Ave. Dur.
INTC	-0.0462	2.3873	-0.1157	0.28	2.04	34.85	0.32
MSFT	0.0108	1.7308	-0.0603	0.53	3.40	32.13	0.33
PFE	-0.0979	1.5244	-0.1475	0.19	1.51	24.49	2.02

Reported are the descriptive statistics of the daily returns of INTC, MSFT and PFE. LB(20) denotes the Ljung–Box statistic on serial correlation of up to 22 lags. The last column reports the average duration between trades (measured in seconds).

Realized Covariance

Let us now turn to the analysis of the three S&P500 component stocks. Table 2.2 presents the descriptive statistics of the daily return series of Intel (INTC), Microsoft (MSFT) and Pfizer (PFE) along with their average duration between trades, i.e. the mean time between two consecutive (valid) transactions measured in seconds. The returns obviously exhibit the well–known characteristics, i.e. a skewed and fat–tailed unconditional distribution.[8] The volatility clustering is nicely illustrated in the time evolvement of the return series and the corresponding realized variances depicted in Figures 2.7 and 2.9, respectively.

Table 2.3 reports some descriptive statistics of the realized volatilities and correlations. In particular, for the latter we follow Barndorff-Nielsen and Shephard (2004a) and compute the daily realized correlation between asset 1 and asset 2 by

$$RCorr_{(1,2),t} = \frac{\sum_{j=1}^{M} r_{(1)j,t} r_{(2)j,t}}{\sqrt{\sum_{j=1}^{M} r_{(1)j,t}^2 \sum_{k=1}^{M} r_{(2)k,t}^2}}. \tag{2.28}$$

Notably, the realized correlations of the three assets considered here all have about the same mean and variance. Moreover, the realized volatilities of the individual stocks are highly serially correlated as is indicated by the Ljung–Box statistics. Interestingly, such pattern is also found for the realized correlations as is illustrated in Figure 2.8 showing significant autocorrelation coefficients of up to roughly half a year (when compared to the Bartlett 95% confidence bands). Figure 2.9 depicts the time evolvement of the realized volatilities and realized correlations, respectively. Apparently, the three assets move together. However, the series are much more erratic than the corresponding realized volatility series of the S&P500 index futures presented in the previous subsection. This may be caused by the use of a lower sampling frequency when constructing the realized covariance matrix.

Our descriptive analysis, obviously, has reproduced the most popular stylized facts of stock returns. However, considering the non–parametric volatility measures,

[8] Skewness and Kurtosis are much less pronounced than those of the S&P500 index futures returns, which, however, are based on a much larger sample period including the crash in 1987 — the so–called "Black Monday".

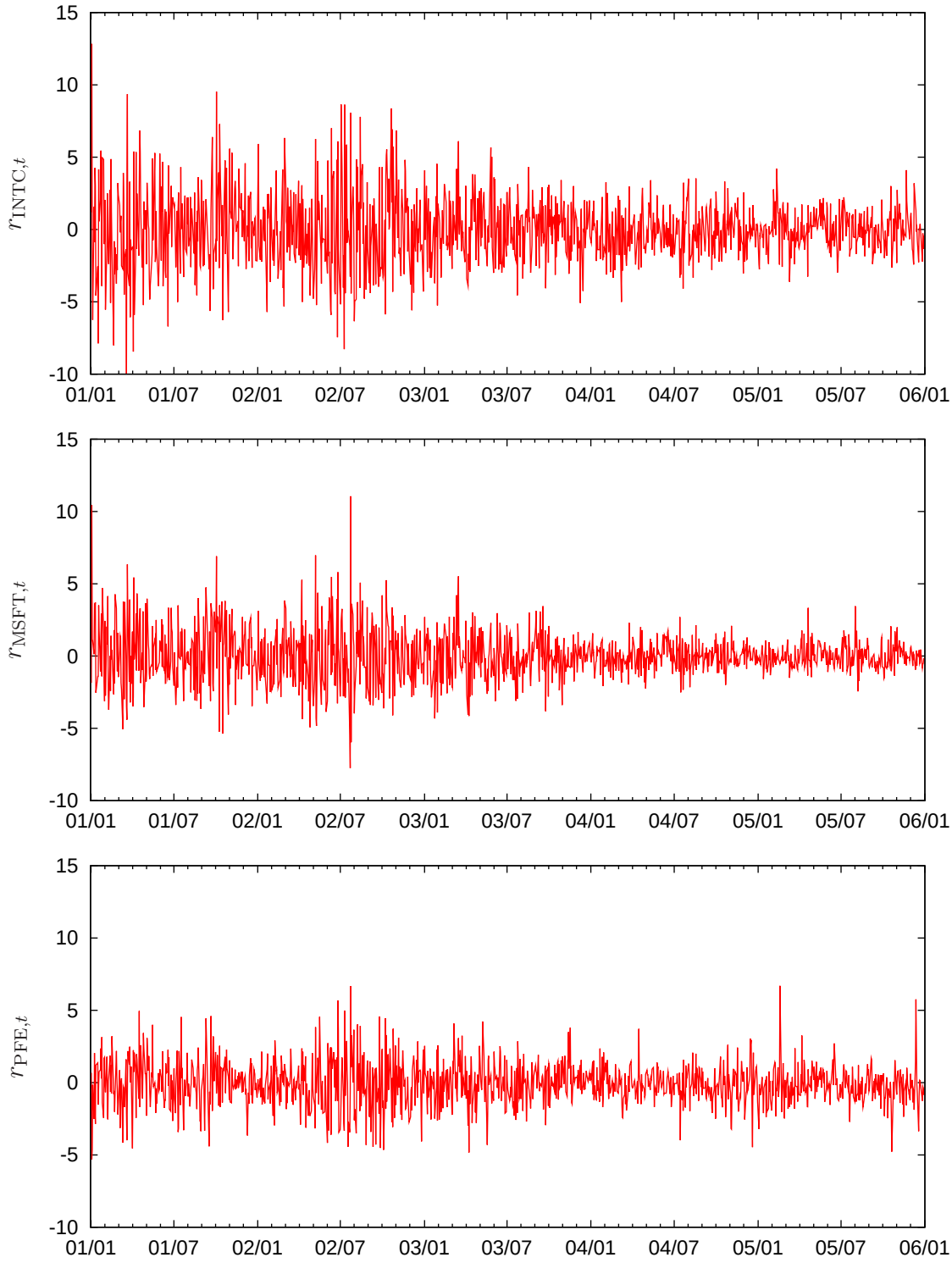

Figure 2.7: Time evolvement of the INTC, MSFT and PFE returns.

Table 2.3: Descriptive Statistics of the Realized Volatilities and Correlations of INTC, MSFT and PFE

Series	Mean			Std.Dev.			Ljung–Box(22)		
	INTC	MSFT	PFE	INTC	MSFT	PFE	INTC	MSFT	PFE
INTC	1.6916	0.2940	0.3351	0.7473	0.3335	0.3196	3,129.1	105.8	166.3
MSFT		1.1611	0.3406		0.5508	0.3273		4,974.8	319.4
PFE			1.4101			0.9263			8,162.1

The diagonal entries report the descriptive statistics of the realized volatilities of the respective assets, whereas the off–diagonal entries report those of the realized correlations as defined in equation (2.28).

i.e. exploiting the information inherent in the high–frequency data, we could also obtain some new insights on the volatility process. In the subsequent chapters of this thesis we make use of this new information and propose three different models for the return and volatility dynamics with each addressing specific aspects of their empirical properties. In particular, the next chapter primarily focuses on the relevance of the volatility of realized volatility in modeling and forecasting. Chapter 4 instead aims at modeling jointly the dynamics and interrelationships of returns, the continuous volatility component and the jump measure. In Chapter 5 the co–movements of multiple assets are modeled via a stochastic volatility model with dynamic conditional correlations, and that additionally accounts for the fact that in practice the realized (co)variance is an unbiased but noisy estimator of the "true" covariance matrix.

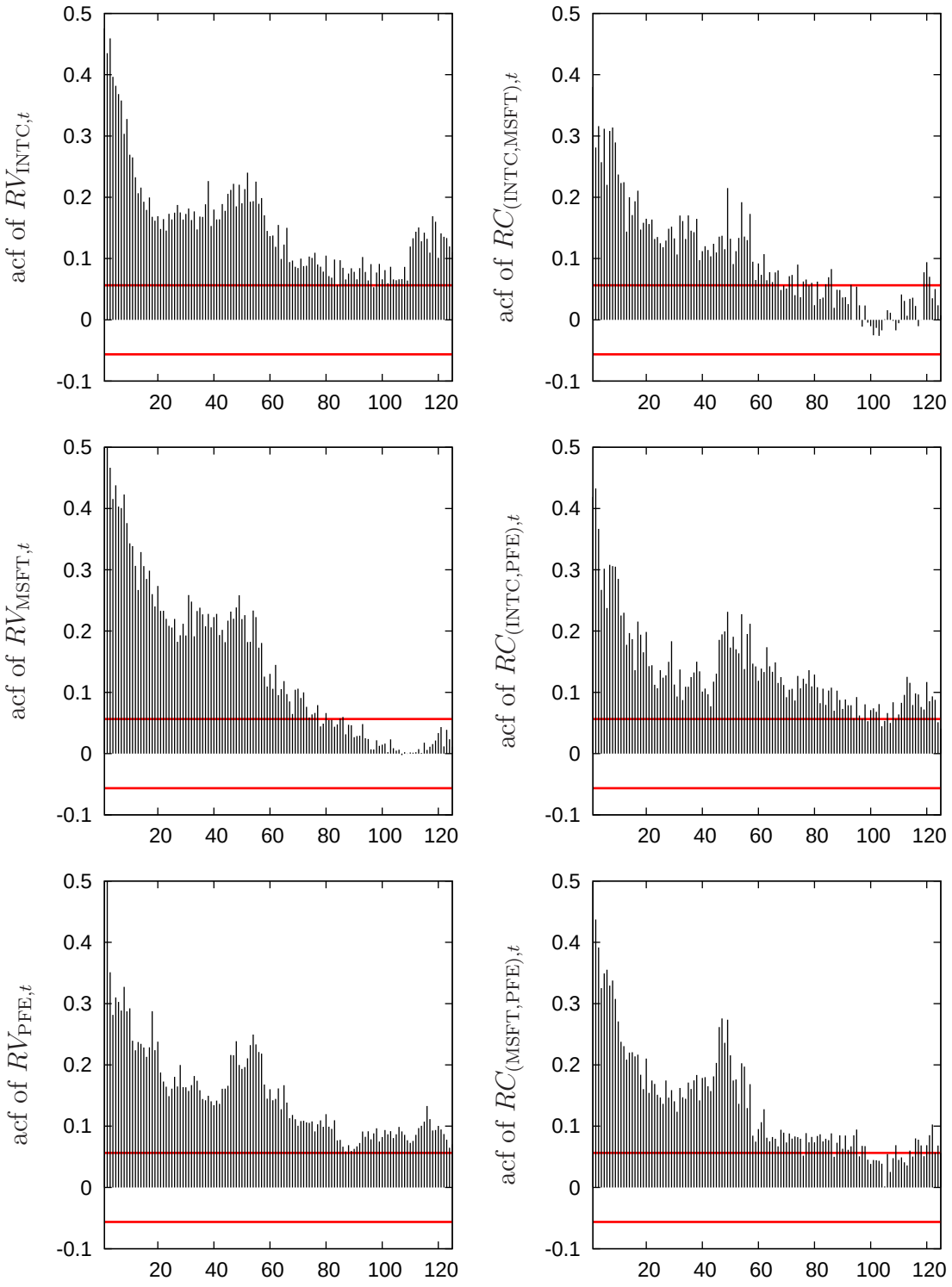

Figure 2.8: Sample autocorrelations of the realized variances and realized correlations (for INTC, MSFT and PFE).

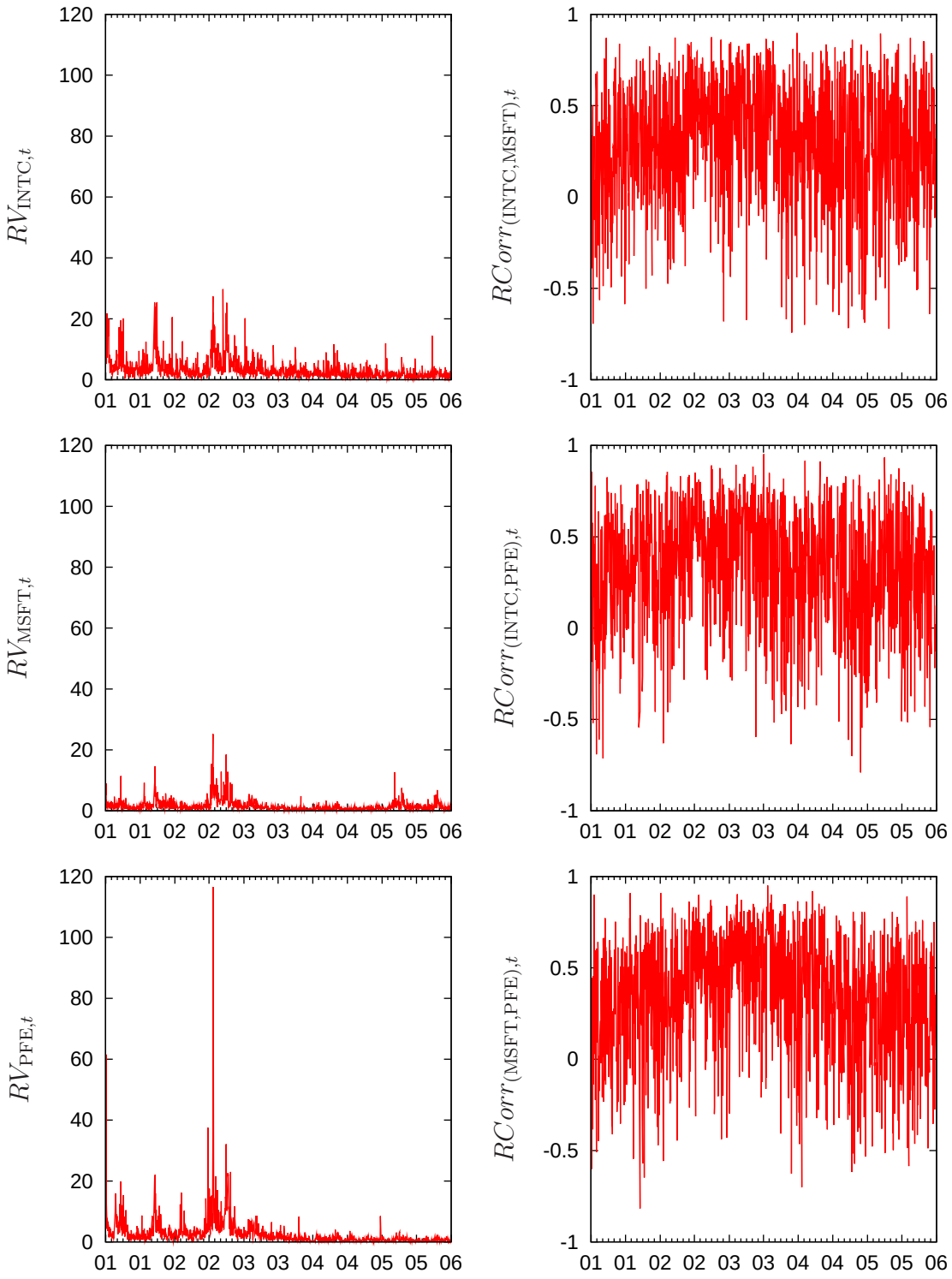

Figure 2.9: Time evolvement of realized variances and realized correlations (for INTC, MSFT and PFE).

3 The Relevance of the Volatility of Realized Volatility for Modeling and Forecasting Volatility

The availability of the non–parametric realized variation measures allows to treat volatility as an "observed" rather than latent variable leading to the development of a series of new and simple–to–implement volatility forecasting models in which the realized volatility is modeled by standard time series procedures. In fact, such reduced–form models for realized volatility have already been considered for a variety of different markets and data sets. The next section provides a review of the Autoregressive Fractional Integrated Moving Average (ARFIMA) and the Heterogeneous Autoregressive (HAR) models, which are the most popular realized volatility models. Both models have been shown to be able to adequately reproduce the observed volatility persistence. Moreover, they generally lead to significant improvements in volatility forecasts relative to conventional stochastic–volatility or GARCH models.

In the ARFIMA as well as in the HAR models it is commonly assumed that the innovations are Gaussian as well as identically and independently distributed (i.i.d.) — an assumption that might be questionable in view of our findings based on the descriptive analysis in Section 2.2.2. In particular, the observed clustering in the volatility of the realized volatility estimator may lead to similar patterns in the errors of the model. Moreover, based on the observed strong skewness and fat–tailedness of the distribution of the realized volatility (and its logarithmic transformation) a more flexible, non–Gaussian distribution might be more adequate. The next section therefore conducts a more thorough residual analysis of the ARFIMA and HAR models. The results indeed reveal that the residuals exhibit volatility clustering and that the Gaussianity assumption is particularly inadequate when modeling the level of the realized volatility. Notably, for the logarithmic realized volatility the skewness and leptokurtosis of the empirical distribution of the ARFIMA and HAR residuals are much less pronounced.

Ignoring the observed findings will lead to inefficiencies when estimating these realized volatility models and result in an inferior forecasting performance. More importantly, in practical applications the presence of time–varying and non–Gaussian conditional distributions can distort risk assessment and, thus, impair risk management.

In this chapter, we therefore investigate the importance of the volatility of real-

ized volatility in modeling and forecasting applications, and propose two extensions of standard realized volatility models. In particular, we allow for non–Gaussian innovations and adopt the more flexible normal inverse Gaussian distribution. Furthermore, to model time–dependent conditional heteroskedasticity we also specify a GARCH specification, which can account for clustering and—to some extent—for the observed unconditional kurtosis. By doing so, we explicitly model the volatility of realized volatility which, to our knowledge, has not yet been considered in the literature.

Our assessment is twofold. Since the standard deviation of realized variance is the main variable of interest for financial applications, our assessment is primarily conducted in terms of realized volatility. In particular, we directly model realized volatility. However, it is also widely accepted that the unconditional distribution of the logarithmic transformation of realized variance is closer to Gaussianity (see, for example, Andersen et al. (2001a,b), and Gonccalves and Meddahi (2005)) leading many researchers to formulate their volatility models in terms of the logarithmic transform. We therefore also consider logarithmic realized variance models. This allows us to investigate the adequacy of the Gaussianity assumption for these models as well as the relevance of the time–variation of the volatility which we also observe for the logarithmic realized variance. The predictive performance of the different models, however, is assessed in terms of realized volatility, which is more relevant from the viewpoint of financial economics.

The remainder of this chapter is organized as follows. The next section provides a brief review and the estimation results of ARFIMA and HAR models for the realized volatility of the S&P500 index futures. Section 3.2 discusses our model extensions and presents the corresponding in–sample estimation results. In Section 3.3 we perform a simulation study to assess the efficiency implications of the proposed extensions. Section 3.4 provides an evaluation of the out–of–sample point and density forecast ability of the different model specifications. Section 3.5 concludes this chapter.

3.1 Popular Realized Volatility Models

Our data analysis and the related empirical literature suggest that the persistence of realized volatility is a distinct feature a realized volatility model should capture. As pointed out earlier, this finding is not only peculiar for the realized volatility measure. In fact, using different volatility proxies numerous empirical studies—starting with Andersen and Bollerslev (1997), Ding et al. (1993), Ding and Granger (1996), and Granger et al. (2000)—show the existence of long memory in volatility. Although at the empirical level the evidence of a strong volatility persistence has been unanimously recognized, at the theoretical level there is much less consensus on the mechanism generating this phenomenon. This is due to the fact that alternative long–memory models are consistent with the data and, thus, empirically indistin-

guishable.[1] As a consequence, the source of long memory in realized volatility is still an open issue. We therefore focus our attention on models that are commonly employed in the existing realized volatility literature.

To capture the long memory in realized volatility or logarithmic realized variance, Andersen et al. (2003) specify the autoregressive fractionally integrated moving average, in short, ARFIMA(p, d, q) model

$$\phi(L)(1-L)^d(y_t-\mu)=\psi(L)u_t, \tag{3.1}$$

with d denoting the fractional difference parameter, $\phi(L)=1-\phi_1 L-\ldots-\phi_p L^p$ and $\psi(L)=1+\psi_1 L+\ldots\psi_q L^q$. Typically, u_t is assumed to be a Gaussian white noise process, and y_t denotes either the realized variance (see, for example, Koopman et al. (2005) or Oomen (2004)) or its logarithmic transform (as first advocated in Andersen et al. (2003)). Retaining the Gaussianity assumption (regardless of the transformation of realized volatility considered), several papers have adopted and extended this model by including, for example, leverage effects (or other nonlinearities) and exogenous variables. The results reported in the literature for different markets and data sets show significant improvements in the point forecasts of volatility when using ARFIMA rather than GARCH–type models.[2] In the context of interval forecasting the distributional assumptions for the error terms should, however, be important.

An alternative model for the realized volatility has been suggested by Corsi (2004). Motivated by the heterogeneous ARCH model of Müller et al. (1997), he proposed the heterogeneous autoregressive (HAR) model, in which the realized volatility is modeled by the sum of (a small number of) volatility components constructed over different time horizons. Although, as such the HAR model is formally not a long–memory model, it has been shown to be able to reproduce quite adequately the hyperbolic decay in the autocorrelation function of realized volatility.

[1] In fact, there exists a large number of different approaches to explain long memory. Historically, the first class of long–memory models has been the fractionally integrated process proposed by Granger and Joyeux (1980) and Hosking (1981) (for comprehensive surveys see Beran (1994) and Robinson (2003)). With another seminal paper showing the link between long memory and the aggregation of an infinite number of stationary processes, Granger (1980) also started an alternative strand of literature, which tries to approximate long–memory dependence through a multi-component approach, as in Andersen and Bollerslev (1997), Engle and Lee (1999), Lux and Marchesi (1999), and Müller et al. (1997). A profoundly different view on the source of long memory is instead offered by, among others, Diebold and Inoue (2001), Gourieroux and Jasiak (2001), Granger and Hyung (2004), Granger and Teräsvirta (1999), and Mikosch and Stărică (2004), who provide theoretical justification and Monte Carlo evidence that models with structural breaks and regime–shifting may exhibit spurious long memory. In addition, other approaches for reproducing long–memory dependence, such as the multifractals and cascade models of Calvet and Fisher (2002, 2004) and Mandelbrot et al. (1997), or the error duration model of Parke (1999), have been proposed (see also Banerjee and Urga (2005) and Davidson and Teräsvirta (2002) for recent reviews on long–memory models).

[2] See, for example, Andersen et al. (2003), Koopman et al. (2005), Martens et al. (2004), Martens and Zein (2004), Oomen (2004), Pong et al. (2004), and Thomakos and Wang (2003), among others.

Defining the k–period realized volatility component by the sum of the single–period realized volatilities, i.e.,[3]

$$\left(\sqrt{RV}\right)_{t+1-k:t} = \frac{1}{k}\sum_{j=1}^{k}\sqrt{RV_{t-j}},$$

the *heterogeneous autoregressive* (HAR) model of Corsi (2004), including the daily, weekly and monthly realized volatility components, is given by

$$\sqrt{RV}_t = \alpha_0 + \alpha_d\sqrt{RV}_{t-1} + \alpha_w\left(\sqrt{RV}\right)_{t-5:t-1} + \alpha_m\left(\sqrt{RV}\right)_{t-22:t-1} + u_t. \quad (3.2)$$

In Corsi (2004), u_t is also assumed to be Gaussian white noise. Employing the volatility–component structure (3.2), simulations reported in Corsi (2004) show that the HAR model is able to reproduce the observed hyperbolic decay of the sample autocorrelations of realized volatility. Moreover, the HAR model's in– and out–of–sample performance is strong and even slightly better than that of an ARFIMA model for realized volatility. A good predictive performance has also been reported in Andersen et al. (2007), who extend the HAR model by including different jump measures. They also consider the HAR model for the logarithm of realized variance, yielding similar results. The logarithmic version of the HAR model considered here is given by

$$\log RV_t = \alpha_0 + \alpha_d \log RV_{t-1} + \alpha_w \left(\log RV\right)_{t-5:t-1} + \alpha_m \left(\log RV\right)_{t-22:t-1} + u_t \quad (3.3)$$

with the multiperiod logarithmic realized–variance components defined by

$$\left(\log RV\right)_{t+1-k:t} = \frac{1}{k}\sum_{j=1}^{k}\log RV_{t-j}.$$

Note, that we follow the above representation, i.e., we formulate the HAR model as a restricted AR(22) model.[4] Hence, the HAR model is nested within the general ARFIMA class, and (in–sample) model comparison is therefore generally straightforward.

In view of the similar performance of ARFIMA and HAR models and given the straightforward estimation of the latter, the HAR model might be preferable in practice. In fact, relative to the HAR model the estimation of the ARFIMA model is non–trivial. The simplest approach is to first estimate the fractional difference parameter, using, for example, the semi–parametric estimator of Geweke

[3] Note that based on Jensen's inequality the volatility components cannot exactly be interpreted as the realized volatility over the specific time interval. However, our definition allows to interpret the HAR model as a restricted AR(22) model. Also, when employing the "true" daily, weekly and monthly realized volatilities—as defined by the square root of the sum of the realized variances—we obtain similar empirical results.

[4] As a consequence the logarithmic realized–variance components cannot be directly interpreted as the logarithm of the multiperiod realized variance.

Table 3.1: Estimation Results for the ARFIMA and HAR Models

Model	Parameter Estimates									AIC	BIC
ARFIMA(0,d,3)	d	ψ_1	ψ_2	ψ_3							
	0.3483 (0.0217)	0.09389 (0.0258)	0.1798 (0.0181)	0.0452 (0.0174)						5484	5517
HAR Model	Mean Eq.				Distribution		Variance Eq.				
	α_0	α_d	α_w	α_m	α	β	ω	α_1	β_1		
I_S	0.1066 (0.0198)	0.4983 (0.0015)	0.2132 (0.0059)	0.1659 (0.0191)			0.1985 (0.0003)			5702	5735
II_S	0.0657 (0.0068)	0.2339 (0.0170)	0.4541 (0.0189)	0.2130 (0.0177)			0.0040 (0.0002)	0.7464 (0.0053)	0.2428 (0.0066)	-67	-21
III_S	0.2180 (0.0075)	0.2540 (0.0068)	0.2285 (0.0077)	0.2645 (0.0078)	1.0313 (0.0498)	0.6740 (0.0479)	0.0933 (0.0034)			-1306	-1260
IV_S	0.0868 (0.0073)	0.2322 (0.0142)	0.3965 (0.0227)	0.2565 (0.0184)	1.6918 (0.1088)	1.054 (0.0975)	0.0034 (0.0003)	0.8143 (0.0117)	0.1237 (0.0110)	-2316	-2257
I_L	−0.0297 (0.0093)	0.3483 (0.0120)	0.3637 (0.0227)	0.2318 (0.0208)			0.2604 (0.0035)			7496	7528
II_L	−0.0386 (0.0090)	0.3089 (0.0187)	0.4005 (0.0282)	0.2331 (0.0234)			0.0430 (0.0050)	0.7369 (0.0241)	0.0963 (0.0083)	7239	7284
III_L	−0.0350 (0.0084)	0.2997 (0.0139)	0.3713 (0.0229)	0.2626 (0.0198)	1.4706 (0.0879)	0.4150 (0.0564)	0.2557 (0.0068)			7018	7063
IV_L	−0.0377 (0.0085)	0.2938 (0.0170)	0.4000 (0.0268)	0.2469 (0.0218)	1.6499 (0.1076)	0.4478 (0.0648)	0.0040 (0.0081)	0.7611 (0.0401)	0.0798 (0.0132)	6899	6958

The different HAR-model specifications are as follows: I is a standard HAR model with Gaussian innovations; II also includes GARCH effects; III is a standard HAR model with (standardized) NIG innovations; and IV corresponds to the HAR–GARCH model with (standardized) NIG innovations. The indices S and L denote the HAR models formulated for realized volatility and logarithmic realized variance, respectively. The numbers in parentheses are the standard errors. The AIC$=-2l+k$, and BIC$=-2l+k\log T$, where l denotes the log likelihood, k the number of parameters in the model, and T is the number of observations.

and Porter-Hudak (1983), and then fit an ARMA model to the filtered series. However, the joint estimation of the ARMA parameters and the fractional difference parameter has been shown to generally improve the accuracy of the estimate of d,[5] though complicating the estimation since the long–memory autocovariance matrix needs to be estimated. Well–known methods for a joint maximum–likelihood estimation of ARFIMA parameters include the approaches of Hosking (1981) and Sowell (1992).[6]

The ARFIMA parameter estimates reported in Table 3.1 are jointly estimated using exact maximum–likelihood with the Geweke–Porter–Hudak estimate serving as the starting value. The AIC and BIC criteria as well as the correlograms of the residuals suggest an ARFIMA(0,d,3) model for the realized volatility of S&P500 index futures. Table 3.1 presents the parameter estimates of the ARFIMA and the standard HAR models, with the latter also being estimated via maximum likelihood.

Figures 3.1 and 3.2 show the results of the residual analysis for the two models. The time series plots and the sample autocorrelation and partial autocorrelation functions of the squared residuals clearly illustrate that the residuals of both models exhibit volatility clustering. In both cases, ARCH–LM tests indicate strong autoregressive conditional heteroskedasticity. This is in line with the time–variation observed in the measures of the volatility of the realized volatility error. Moreover, the QQ–Plot and the kernel density estimates in Figures 3.1 and 3.2 convincingly illustrate the inadequacy of the normality assumption for both models. Note that for the logarithmic realized variance we find the same form of time–dependent heteroskedasticity for both models. However, the volatility clustering is less pronounced, a finding that is consistent with the characteristics of the measures for the volatility of logarithmic realized variance. Also, as expected, the non–Gaussianity is less pronounced, but still existent.

[5] Agiakloglou et al. (1993) report poor small–sample properties of the Geweke–Porter–Hudak estimator.

[6] See Doornik and Ooms (2003) for a recent review on this topic.

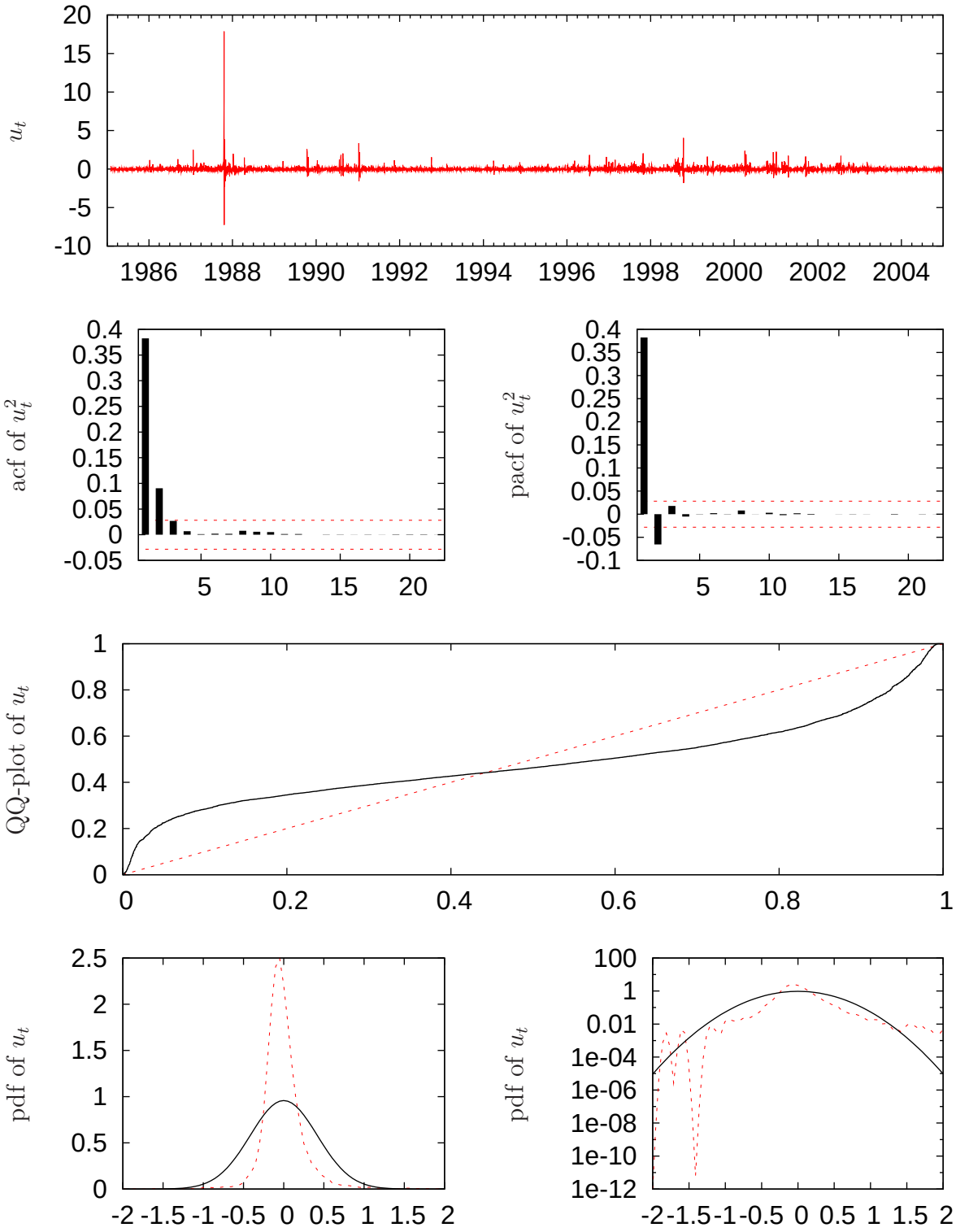

Figure 3.1: Residual analysis of the Gaussian ARFIMA(0,d,3) model for realized volatility. Shown are the time series of the residuals (upper panel), the sample autocorrelation functions (acf) and partial autocorrelation functions (pacf) of the squared residuals (second panel), the quantile–quantile plot (third panel) and on the bottom panel the kernel density estimates of the residuals (dashed line) and the estimated normal density (solid line) in level (left) and log scales (right).

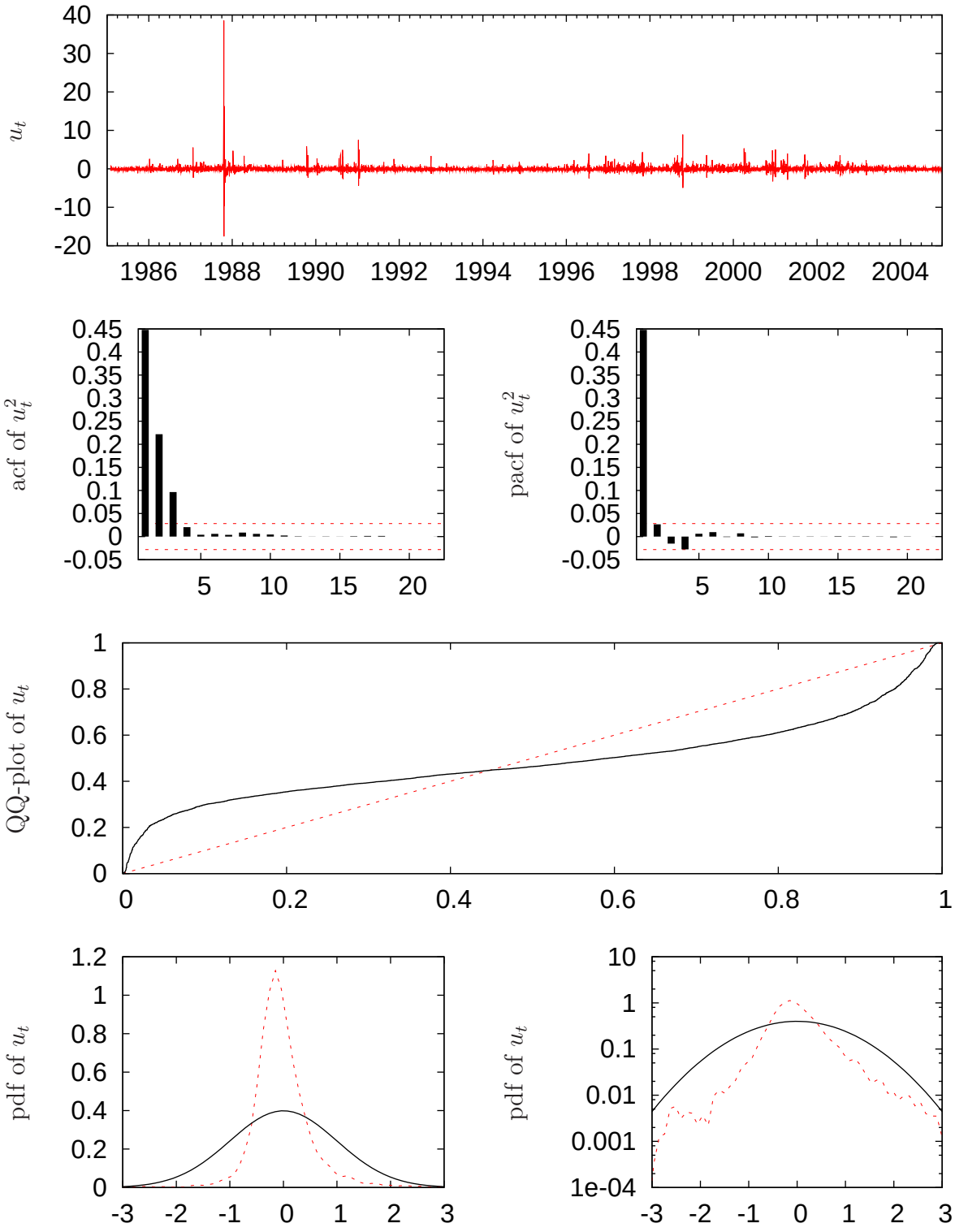

Figure 3.2: Residual analysis of the Gaussian HAR model for realized volatility. Shown are the time series of the residuals (upper panel), the sample autocorrelation functions (acf) and partial autocorrelation functions (pacf) of the squared residuals (second panel), the quantile–quantile plot (third panel) and on the bottom panel the kernel density estimates of the residuals (dashed line) and the estimated normal density (solid line) in level (left) and log scales (right).

3.2 Persistence, Volatility Clustering and Fat Tails

The empirical results presented above point to two major shortcomings of the long memory realized volatility models commonly used. Firstly, the assumption of Gaussian innovations appears inadequate. Secondly, the i.i.d. assumption of the error terms is clearly violated in view of the time–variation and clustering of the residual variance. In the following we therefore propose two model extensions that should account for these properties. Since HAR and ARFIMA models behave similarly in terms of forecasting and model misspecifications, we focus our discussion solely on extended HAR models. The proposed modifications can be straightforwardly adopted in an ARFIMA framework—though the estimation will be even more challenging. We expect the results for the extended HAR and ARFIMA models to be compatible. In fact, in Corsi (2004) it was shown that good fitting as well as good point forecast performance coincide in the HAR model. In particular, the empirical results indicate both, a slightly improved in–sample and out–of–sample performance of the HAR over the ARFIMA model.

To account for the observed volatility clustering in realized volatility, we extend the HAR model by including a GARCH component, giving rise to the HAR–GARCH(p,q) model

$$\begin{aligned} y_t &= \alpha_0 + \alpha_d y_{t-1} + \alpha_w \left(y\right)_{t-5:t-1} + \alpha_m \left(y\right)_{t-22:t-1} + \sqrt{h_t} u_t \\ h_t &= \omega + \sum_{j=1}^{q} \alpha_j u_{t-j}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \\ u_t|\Omega_{t-1} &\sim (0,1), \end{aligned} \tag{3.4}$$

where Ω_{t-1} denotes the σ-field generated by all the information available up to time $t-1$, and y is either $\sqrt{RV}$ or $\log RV$. The error term, $\sqrt{h_t}u_t$, follows a conditional density with time–varying variance.

Although the incorporation of the GARCH specification can produce fatter unconditional tails, the normality assumption does not allow for the observed skewness. To deal with the non–Gaussianity of the error terms we specify a standardized normal inverse Gaussian (NIG) distribution for the (unconditional) i.i.d. innovations u_t. The NIG distribution, introduced by Barndorff-Nielsen (1997, 1998), is rather flexible and able to reproduce a range of symmetric and asymmetric distributions, such as the normal. Its density is given by

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

where $K_i\left(x\right)$ is the modified Bessel function of the third kind[7] and index 1; $\mu \in \mathbb{R}$ denotes the location parameter, $\delta > 0$ the scale, $\alpha > 0$ and $\beta \in (-\alpha, \alpha)$ the shape

[7] Note, that this function is also oftentimes called the modified Bessel function of the third kind or Macdonald function.

parameters, with $\beta = 0$ indicating a symmetric distribution. Mean and variance are given by

$$\mathrm{E}[x] = \mu + \frac{\delta\beta}{\sqrt{\alpha^2 - \beta^2}} \qquad \text{and} \qquad \mathrm{Var}[x] = \frac{\delta\alpha^2}{\sqrt{\alpha^2 - \beta^2}^3}.$$

To derive the *standardized* NIG distribution with zero mean and unit variance, we solve the resulting equations and derive the values of μ and δ in terms of α and β, and thus obtain

$$\mu = -\frac{\beta\left(\alpha^2 - \beta^2\right)}{\alpha^2} \qquad \text{and} \qquad \delta = \frac{\left(\alpha^2 - \beta^2\right)^{3/2}}{\alpha^2}.$$

By combining the HAR model with a standardized NIG distribution and a GARCH specification, we obtain a quasi–long–memory model that should be able to capture both non–Gaussianity and time–dependent conditional heteroskedasticity.

Note that these extensions can also be adopted in the ARFIMA framework and have, in part, been considered in Baillie et al. (1996), who propose an ARFIMA–GARCH model to analyze inflation.

Maximum–likelihood estimates for various HAR specifications are presented in Table 3.1. Specifically, we extend the conventional HAR model with Gaussian innovations (Model I) by including a GARCH(1,1) specification (Model II); Model III corresponds to Model I but with zero mean NIG–distributed errors; and Model IV includes both modifications, i.e. we allow for NIG–distributed innovations and conditional heteroskedasticity. The corresponding results of the HAR model formulated for realized volatility are indexed by S, those for logarithmic realized variance are indexed by L.

The results show that the GARCH extension substantially improves the goodness of fit, as measured by the AIC and BIC criteria, especially for realized volatility. Both criteria as well as the ARCH–LM test suggest a GARCH(1,1) specification, which is also the preferred choice when modeling the volatility of asset returns. Comparing the parameter estimates of the mean equation of Model II to those of the standard HAR Model I, we observe an increase in the parameter of the weekly volatility component, α_w, while, at the same time, the influence of realized volatility lagged by one day decreases when including the GARCH specification.

It is well–known that for a GARCH process the kurtosis of the dependent variable is determined by both the kurtosis of the error distribution and the persistence in the GARCH equation, i.e., by $\alpha_1 + \beta_1$ in a GARCH(1,1) process. Bai et al. (2003) have shown that the commonly reported parameter estimates, which are in the range of $0.85 < \hat{\alpha}_1 + \hat{\beta}_1 < 1$, are not sufficient for generating the observed kurtosis when assuming normally distributed errors. Given that our persistence estimate under the Gaussianity assumption lies within the (0.83, 1)–interval and that the kurtosis of realized volatility is much stronger than is commonly found for asset returns, simply adding a GARCH specification will not suffice to capture the observed kurtosis of realized volatility. A more heavy–tailed distribution for

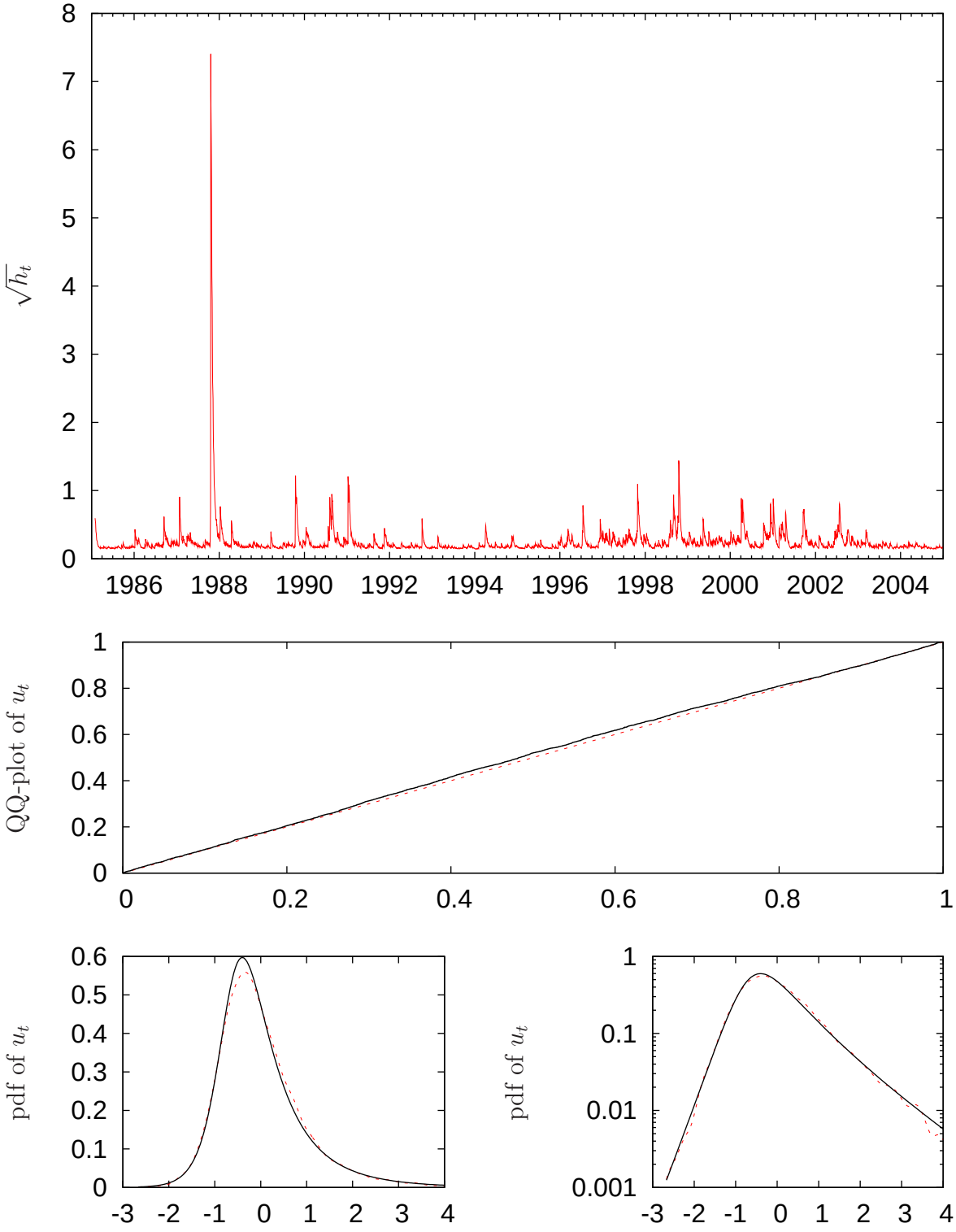

Figure 3.3: Diagnostics of the HAR–GARCH(1,1)–NIG model for realized volatility. The upper panel presents the GARCH–filtered volatility of realized volatility series; the middle panel shows the quantile–quantile plot of the residuals, the bottom panel presents the kernel density estimates of the residuals (dashed line) and the estimated NIG density (solid line) in level (left) and log scales (right).

the innovations is required. Given that the logarithmic transformation of realized variance reduces skewness and kurtosis, the need for a more flexible distribution might be less important in this case than for the realized volatility measure.

According to the goodness–of–fit measures reported in Table 3.1 replacing the Gaussian by the NIG distribution greatly improves the models' fit both for the conventional and the GARCH specification (Models I and II)—with the improvement being again more pronounced for realized volatility than for the logarithmic transform. For both measures the overall preferred model turns out to be the HAR–GARCH specification with NIG–distributed innovations, suggesting that both extensions to the standard long–memory model are important. This finding is in line with the GARCH literature for asset returns. For example, Verhoeven and McAleer (2004) and Mittnik et al. (1998, 2000) show that GARCH models with skewed and leptokurtic errors outperform their Gaussian counterparts.

Introducing the NIG distribution affects the GARCH–parameter estimates. For realized volatility the persistence is much less than under the Gaussian assumption since excess kurtosis can, in part, be captured by the shape parameter of the NIG distribution, α. Note also that, in comparison to the HAR–NIG model without GARCH specification, the shape parameter is much larger, indicating less kurtosis.

Figure 3.3 demonstrates the adequacy of the HAR–GARCH(1,1)–NIG model for realized volatility. Both skewness and the tail behavior of the innovations are well captured by the NIG. The GARCH–filtered volatility series (top panel in Figure 3.3) shows the clustering in the volatility of realized volatility. In fact, the time series pattern of the filtered series is, though less pronounced, similar to the characteristics observed in the measures of the volatility of realized volatility discussed in Section 2.2.2. We find similar but less pronounced results for the logarithmic realized variance models.

3.3 Gains in Efficiency

Ignoring the presence of heteroskedasticity and non–Gaussianity in innovations leads to inefficient parameter estimates when estimating the (standard) HAR and ARFIMA models proposed in the literature. Inaccurate estimates do not only hamper their interpretation but also affect forecasting accuracy.

To assess the effects of explicitly allowing for heteroskedasticity and non–Gaussianity on efficiency, we conduct a simulation study. We generate 1,000 series, each with sample size 5,000, from a HAR–GARCH(1,1) model with standardized NIG–distributed innovations, using the estimates reported in Table 3.1. For each replication we consider the first 500, 1,250, 2,500, and, finally, all 5,000 data points which correspond to about 2, 5, 10 and 20 years of daily data, respectively, with the latter approximating the sample size of our data set. From the simulated data we estimate the HAR specifications I-IV discussed in the previous section.

Table 3.2 reports the root mean square error of the parameter estimates for the four models for realized volatility from the 1,000 simulation runs. Focusing first

Table 3.2: Efficiency Results for Realized Volatility Models

Obs.	Model	Mean Eq.				Distribution		Variance Eq.		
		α_0	α_d	α_w	α_m	α	β	ω	α_1	β_1
500	I_S	0.0763	0.0818	0.1403	0.1238			0.0606		
	II_S	0.0723	0.0672	0.1216	0.1185			0.0061	0.1953	0.1055
	III_S	0.0940	0.0536	0.1245	0.1029	0.5634	0.4313	0.0650		
	IV_S	0.0493	0.0455	0.0825	0.0816	0.5355	0.4272	0.0031	0.1017	0.0471
1250	I_S	0.0392	0.0592	0.0952	0.0794			0.0589		
	II_S	0.0338	0.0428	0.0722	0.0679			0.0023	0.0802	0.0508
	III_S	0.0652	0.0358	0.0955	0.0660	0.5945	0.4501	0.0558		
	IV_S	0.0229	0.0278	0.0491	0.0470	0.2376	0.2032	0.0011	0.0383	0.0276
2500	I_S	0.0256	0.0409	0.0650	0.0565			0.0571		
	II_S	0.0223	0.0302	0.0499	0.0476			0.0013	0.0501	0.0370
	III_S	0.0559	0.0285	0.0837	0.0481	0.6082	0.4607	0.0535		
	IV_S	0.0149	0.0203	0.0333	0.0310	0.1466	0.1262	0.0007	0.0252	0.0195
5000	I_S	0.0179	0.0343	0.0537	0.0438			0.0529		
	II_S	0.0141	0.0219	0.0361	0.0335			0.0008	0.0325	0.0251
	III_S	0.0504	0.0223	0.0775	0.0387	0.6270	0.4743	0.0513		
	IV_S	0.0099	0.0139	0.0237	0.0220	0.1112	0.0938	0.0004	0.0166	0.0131

All entries report root mean square error of parameter estimates for the different models. They are based on 1,000 simulations from the HAR–GARCH–NIG model as given in Table 3.1. "Obs." denotes the number of simulated observations of each simulation run and "Model" corresponds to the different models: I is a standard HAR model with Gaussian innovations; II also includes GARCH effects; III is a standard HAR model with (standardized) NIG innovations; and IV corresponds to the HAR–GARCH model with (standardized) NIG innovations.

on the results for the parameters of the HAR mean equation, we see that the inclusion of the GARCH specification yields greater improvements in efficiency than just allowing for NIG–distributed innovations, indicating that the incorporation of conditional heteroskedasticity is more relevant. Notably, although the HAR model with NIG errors is generally more efficient than the standard HAR specification, it has difficulties in estimating the constant and the parameter of the weekly volatility component. This is in line with our discussion in the previous section. However, when allowing for a GARCH specification, these problems vanish. As expected, given that it matches the data generation process, the HAR–GARCH–NIG model exhibits the strongest gains in parameter efficiency. The results suggest that allowing for volatility clustering leads to substantial efficiency gains. The implications of the NIG extension on efficiency is somewhat ambiguous. The forecasting experiment reported in the next section will shed additional light on this issue.

Turning our attention to the parameters of the variance equation in Table 3.2 we find that, for larger sample sizes, the efficiency results for the (wrongly specified) Gaussian HAR–GARCH model are surprisingly good. The results for the parameters of the NIG distribution should not be taken too seriously in the case of the standard HAR model, given the trade–off between the kurtosis induced by the GARCH specification and by the NIG distribution. However, for the HAR–GARCH specification we can conclude that a sufficiently large sample size is required to accurately estimate the distributional parameters. Similar efficiency results are found for the logarithmic realized variance models and are therefore not reported.

3.4 Forecast Evaluation

In order to assess the relevance of allowing for conditional heteroscedasticity of realized volatility for out–of–sample forecasts we consider the period from December 13, 1988 to December 30, 2004 of the futures data, providing us with 4,040 forecasts. We estimate all four model specifications from the first 1,000 observations (January 1, 1985 to December 12, 1988) and recursively construct one–step–ahead realized volatility forecasts. In each recursion we re–estimate, expanding the data set by one observation.

Since the standard deviation of realized variance is the relevant variable in many applications, e.g. for risk assessment and management applications, we evaluate the predictive performance of the different models in terms of their ability to forecast realized volatility. Moreover, to assure comparability in the forecasting performance of the realized volatility/logarithmic realized variance models, we explicitly account for the bias induced by the transformation of the logarithmic realized–variance forecasts. In particular, the realized volatility forecasts of the logarithmic realized variance models are computed as

$$\widehat{\sqrt{RV_{t|t-1}}} = \mathrm{E}(\sqrt{RV_{t|t-1}}) = \exp\left\{\frac{1}{2}\widehat{\log RV_{t|t-1}}\right\} M(\frac{1}{2}h_t), \tag{3.5}$$

with $\log \widehat{RV_{t|t-1}}$ denoting the one–step–ahead logarithmic realized–variance forecast for day t constructed from information available up to time $t-1$, and $M(\cdot)$ denoting the moment generating function of the assumed distribution. For the standardized normal distribution $M(\zeta) = \exp\left\{\frac{\zeta^2}{2}\right\} \quad \forall \zeta \in \mathbb{R}$, and for the standardized NIG distribution

$$M(\zeta) = \exp\left\{\frac{(\alpha^2-\beta^2)}{\alpha^2}\left(-\beta\zeta + (\alpha^2-\beta^2)\left(1 - \frac{\sqrt{\alpha^2-(\beta+\zeta)^2}}{\sqrt{\alpha^2-\beta^2}}\right)\right)\right\},$$

for all $\zeta \in [-(\alpha+\beta), \alpha-\beta]$.

To evaluate the predictive performance of the different volatility models we follow Andersen and Bollerslev (1998) and Andersen et al. (2003), among others, and compute the R^2 statistic from the Mincer–Zarnowitz regressions of observed realized volatility on the corresponding forecasts. In addition to the R^2 statistic, we also report the root mean square forecast error (RMSE), the mean absolute error (MAE) and the root mean squared percentage error (RMSPE). Assigning more weight to large volatility–forecast errors, the RMSE will be of particular interest under a risk–management perspective.[8]

The results of the forecasting exercise are presented in Table 3.3. The first four columns report the criteria for the one–step–ahead point forecast evaluation criteria for the different HAR models. The results show that, regardless of the transformation considered, all four statistics yield more or less the same performance rankings for the four models. The best model for forecasting daily volatility is clearly the HAR model with GARCH specification and Gaussian distributed innovations, and we find the HAR–GARCH–NIG model to perform second best. Surprisingly, the forecasts obtained from the HAR model with NIG assumption perform worst.

Comparing the evaluation criteria across all eight specifications shows that generally the logarithmic realized variance models perform marginally better than their realized volatility counterparts. The Gaussian HAR–GARCH model formulated for logarithmic realized variance provides the best point forecasts according to all criteria, with the single exception of the RMSPE favoring the same model but expressed for realized volatility.

In summary, the forecasting evaluations show that allowing for time–varying volatility of realized volatility/logarithmic realized variance improves the accuracy of volatility point forecasts and should, therefore, be important for risk management applications, such as Value–at–Risk calculations. In contrast, permitting skewness and leptokurtosis in the innovation distribution does not seem to help in point forecasting. This is somehow in line with the conclusions drawn from our efficiency simulations discussed in the previous section. Moreover, forecasting realized volatility based on the logarithmic realized variance models results in an improvement in forecast accuracy.

[8] Note, however, that under the null hypothesis of the Mincer–Zarnowitz test for unbiasedness of forecasts, the RMSE is a homogeneous function of the regression coefficient.

Table 3.3: One–step–ahead Forecast Evaluation

Model	R^2	RMSE	MAE	RMSPE	$S(f_{t\|t-1}, y_t)$
I_S	0.4848	0.3161	0.1926	0.3298	-1725.9
II_S	0.5109	0.3084	0.1849	0.2925	-98.8
III_S	0.5034	0.3173	0.1996	0.3422	538.8
IV_S	0.5074	0.3091	0.1870	0.3034	941.3
I_L	0.4972	0.3108	0.1869	0.3165	865.9
II_L	0.5228	0.3033	0.1822	0.2934	931.5
III_L	0.5091	0.3126	0.1884	0.3272	1022.7
IV_L	0.5169	0.3052	0.1842	0.2961	1068.7

"Model" represents the different model specifications: I is a standard HAR model with Gaussian innovations; II also includes GARCH effects; III is a standard HAR model with (standardized) NIG innovations; and IV corresponds to the HAR–GARCH model with (standardized) NIG innovations. The indices S and L denote the HAR models formulated for realized volatility and logarithmic realized variance, respectively. The reported R^2 are the regression coefficients of realized volatility on a constant and volatility forecasts. $S(f_{t|t-1}, y_t)$ are the logarithmic scores as defined in (3.7) of the one–step–ahead density forecasts.

Although volatility point forecasting is, in general, of primary interest, interval or density prediction can also be of interest. Firstly, when forecasting returns, the uncertainty associated with the volatility estimates carries over to the uncertainty of return forecasts. Secondly, the increasing availability of volatility–based derivatives renders volatility density forecasts to be more and more relevant (see, for example, Corradi et al. (2005, 2006), who develop an estimator of the predictive density of the integrated variance — measured in terms of realized variance).

Because time–variation in the volatility of realized volatility implies time–varying conditional densities, we also evaluate the accuracy of the one–step–ahead density forecasts using a method proposed by Diebold et al. (1998), which is based on Rosenblatt (1952). With $\{f_{t|t-1}(y_t)\}_{t=1}^{n}$ denoting a sequence of n density forecasts constructed from a parametric forecasting model using information available up to time $t-1$, the probability–integral transform z_t is defined as

$$z_t = \int_{-\infty}^{y_t} f_{t|t-1}\left(u\right) du. \tag{3.6}$$

The quantity z_t represents the cumulative distribution function based on the forecasted density evaluated at the realization y_t, i.e., at $\sqrt{RV_t}$. If the density forecast is correct, then the sequence of probability–integral transforms, $\{z_t\}_{t=1}^{n}$, is i.i.d. uniformly distributed on the unit interval (see Diebold et al. (1998)).

Figure 3.4 presents the histograms of the corresponding probability–integral transforms of the four models for realized volatility. It turns out that the z_t of the models with Gaussian innovations are far from being uniformly distributed, although the incorporation of the volatility of realized volatility leads to some improvement. The inability to capture skewness is clearly illustrated by the graphs. In contrast, the NIG–based HAR models provide very accurate density forecasts with the HAR–GARCH specification being somewhat superior. These results strongly favor the NIG extension of HAR and HAR–GARCH models. We find similar results for the probability–integral transforms of the logarithmic realized variance models, which are, however, closer to being uniformly distributed than the corresponding series of the realized volatility models.

To better quantify these visual results, we additionally compute the logarithmic score of the density forecasts defined as (see, for example, Matheson and Winkler (1976))

$$S(\{f_{t|t-1}\}_{t=1}^{n}, \{y_t\}_{t=1}^{n}) = n^{-1} \sum_{t=1}^{n} \log f_{t|t-1}(y_t), \tag{3.7}$$

with $f_{t|t-1}$ denoting again the predicted density, and y_t the realized value of the variable to be forecast. Using the logarithmic scoring rule allows us to explicitly rank the competing density forecasts. A high score, implying a high likelihood for the observed y_t–realization, indicates a better predictive performance.

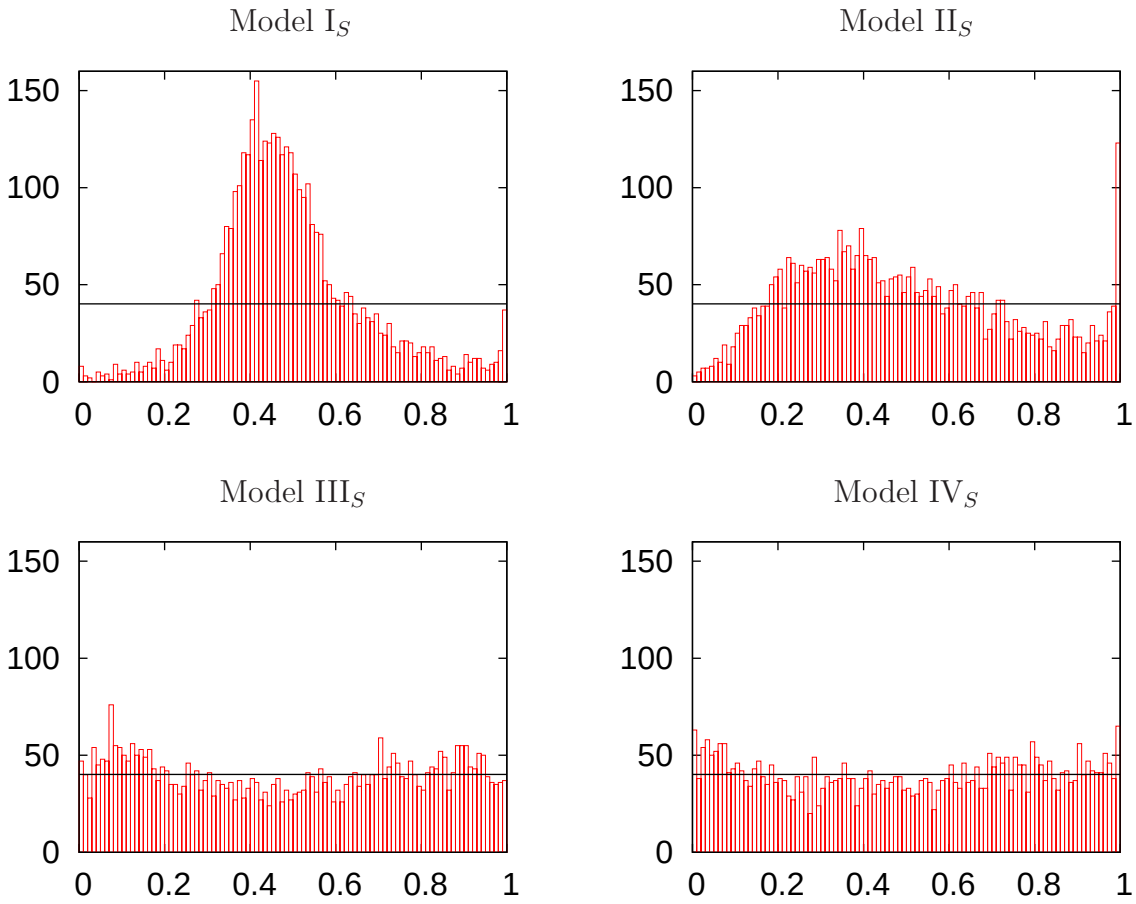

Figure 3.4: Probability–integral transforms of density forecasts based on the four realized volatility models: Model I$_S$ refers to the standard HAR model with Gaussian innovations; Model II$_S$ also includes GARCH effects; Model III$_S$ is a standard HAR model with standardized NIG innovations; and Model IV$_S$ corresponds to the HAR–GARCH(1,1) model with standardized NIG innovations.

The logarithmic scores are presented in the last column of Table 3.3.[9] The results confirm the conclusions drawn from the histograms of the probability–integral transform series in Figure 3.4, i.e. the HAR–GARCH(1,1)–NIG models provide the best density forecasts; and the density forecasts based on the logarithmic realized variance models outperform all of their realized volatility counterparts.

The fact that NIG–based HAR/HAR–GARCH models provide better realized volatility density forecasts, but at the same time perform worse than the Gaussian HAR/HAR–GARCH models in point forecasting may appear counterintuitive. However, it is widely reported in the (point) forecasting literature that more parsimonious, though potentially misspecified models, may generate more accurate point forecasts (see e.g. Clements and Hendry (1998) or Lütkepohl (1993)).[10]

3.5 Conclusion

We have shown that the commonly used reduced–form realized volatility models, such as the ARFIMA or HAR models are better characterized by non–Gaussian innovations and time–varying volatility. Specifically, we favor a heavy–tailed distribution, such as the NIG distribution, and more importantly a model that allows for GARCH–type clustering in the volatility of volatility. In–sample estimation results show an overwhelming superiority of the HAR–GARCH model with NIG distributed innovations. It appears to be important to incorporate the volatility of volatility and to specify an adequate distribution of the error terms, when modeling realized volatility or the logarithm of realized variance. As the time–variation in the innovations might be partly attributed to the time–variation in the volatility of the realized volatility estimator, we expect that the incorporation of this pattern is also relevant for other realized volatility models, such as e.g. the unobserved ARMA component (UC) model of Barndorff-Nielsen and Shephard (2002a), that was also applied in Koopman et al. (2005).

Investigating the implications of the two proposed extensions for the efficiency of the parameter estimates we conclude that the time–varying volatility of realized volatility/logarithmic realized variance is of importance and a GARCH–type extension should be incorporated. Our forecasting experiments suggest that for accurate point forecasts the GARCH specification is more important than allowing for a fat–tailed and possibly skewed distribution. For risk management applications, such as Value–at–Risk calculations, the time–variation of the volatility of volatility should,

[9] Since the density forecasting performance is also evaluated in terms of realized volatility, the logarithmic scores of the logarithmic realized variance models are given by $S(\{f_{t|t-1}\}_{t=1}^{n}, \{\sqrt{RV_t}\}_{t=1}^{n}) = n^{-1} \sum_{t=1}^{n} \log\left(\frac{f_{t|t-1}(\log RV_t)}{2\sqrt{RV_t}}\right)$.

[10] Much less is known about the effects of model parsimony on the accuracy of density forecasts. A small simulation study comparing the point and density forecast accuracy of the pure Gaussian–HAR model and the HAR–NIG model, with the latter being the correct data generating process, corroborate our empirical observation. In particular, we find strong superiority of the HAR–NIG model in density forecasting, but slightly less accurate point forecasts than those produced by the Gaussian–HAR model.

therefore, be incorporated, whereas the specification of a non–Gaussian distribution is not necessarily required. However, whenever interval or density forecasts of realized volatility are the main focus — for example when pricing or assessing the risk of volatility derivatives — a flexible distribution, such as the NIG, should be specified in addition to a GARCH component. Moreover, accounting for the uncertainty associated with the realized volatility forecasts may also improve the accuracy of return forecasts. Our forecasting results also revealed that, regardless of the application at hand, i.e. whether point or density forecasts are of more interest, modeling logarithmic realized variance rather than realized volatility directly is preferable when forecasting realized volatility.

The relevance of our findings for extended realized volatility models, for example models including leverage effects (see Martens et al. (2004)), will be interesting. We expect, however, that, although the incorporation of the leverage effect may lead to some reduction in the skewness of the innovations, the specification of a non–Gaussian distribution might still be necessary in order to reproduce the observed excess kurtosis. We also expect that the time–variation in the volatility of realized volatility will remain, since part of it can be attributed to the volatility of the realized volatility estimator. Note that the variance of this estimator currently gains increasing interest in the literature, see e.g. Barndorff-Nielsen et al. (2006a), who derive more efficient measures of this quantity. Of course, alternatively to our GARCH–specification, these measures may also be utilized to incorporate the volatility of volatility into the realized volatility models. Under a forecasting perspective the specification of a an extra process for this measure may be useful.

4 A Discrete–time Model for Daily S&P500 Returns and Realized Variations: Jumps and Leverage Effects

A current theme in the volatility literature also concerns the question of whether financial prices, and equity prices in particular, may be adequately described by continuous sample path processes, or whether the price movements exhibit discontinuities, i.e. jumps. In addition to the implications for the direct modeling of the price process — as is the focus of this chapter — the answer to this question has important implications for risk management and asset pricing more generally.[1] In the discrete–time framework, however, it is difficult to distinguish between the continuous time evolvement and jumps due to the unobserved price evolvement between two consecutive observations.

One strand of the literature has therefore sought to answer the question through the estimation of specific parametric continuous–time models. This literature dates back to the early work of Merton (1976), with more recent contributions allowing for both jumps and time–varying stochastic volatility including Andersen et al. (2002), Bates (2000), Chernov et al. (2003), Eraker (2004), Eraker et al. (2003), and Pan (2002), among others. Still, the estimation of parametric jump diffusion models remains difficult, and the existing empirical results based on daily or coarser frequency data typically do not allow for a very clear distinction between pure diffusion multi–factor stochastic volatility models and lower–order models with jumps. Of course, given the often large within–day price movements, the daily data most often used in the estimation of the models may simply not be informative enough to provide a firm answer. At the same time, the direct estimation of specific parametric volatility models with large samples of high–frequency intraday data remains extremely challenging from a computational perspective and moreover requires that all of the market microstructure complications inherent in the high–frequency data be properly incorporated into the model.

This in turn has motivated a second more recent strand of the literature in which the intraday data is summarized into the lower–frequency non–parametric daily realized variance and Bipower variation measures, which allow for a direct

[1] E.g. for asset pricing the presence of jumps introduces an additional source of risk that needs to be priced adequately.

non–parametric decomposition of the total price variation into its two separate components. Utilizing these ideas, Andersen et al. (2007) and Huang and Tauchen (2005) both report empirical evidence in support of significant contributions to the overall daily price variation coming from the jump component, which is also supported by our empirical findings in Chapter 2.

The apparent relevance of jumps and the distinctly different distributional and dynamic features of the two volatility components (see Chapter 2) suggests a more structural approach to volatility modeling. At least to our knowledge, none of the existing models using the realized variation measures have pursued this idea. Although, by modeling solely the realized volatility, i.e. total price variation, these reduced–form models disregard potential valuable information inherent in the different volatility measures.

In this chapter we therefore propose a discrete–time stochastic volatility model for returns, Bipower variation and jumps. The explicit decomposition of the two volatility components and the joint modeling with the returns allows us to assess additionally whether the often observed asymmetry between lagged returns and current volatility, oftentimes referred to as the "leverage effect", works through the continuous volatility component and/or the jumps. As other studies (see e.g. Bollerslev et al. (2006b)) also point towards the existence of a contemporaneous leverage effect showing up in probably cross–correlated disturbances of the return and volatility equations, we first estimate our system equation by equation assuming the independence of the disturbances. We then assess the presence and form of the contemporaneous inter–dependencies based on the residuals. The results reveal important nonlinear inter–dependencies among the residuals, and we therefore go on to account for these in a general recursive simultaneous equation system. Despite the general and very flexible structure of our model, maximum–likelihood estimation remains relatively straightforward. Our model estimates are based on daily realized volatilities and returns constructed from high–frequency five–minute S&P500 index futures over the 1985 to 2004 sample period. Moreover, a subsequent simulation study based on parametric bootstrapping also confirms the adequacy of our final preferred model specification.

The remainder of the Chapter is structured as follows. Section 4.1 introduces the three basic equations for returns, Bipower variation and the relative jump series. Section 4.2 presents the equation–by–equation estimation results as well as an assessment of the inter–dependencies among the cross–equation disturbances. Section 4.3 describes the joint recursive model and the corresponding maximum–likelihood estimates. Simulations from the model are used in Section 4.4 to further investigate the adequacy of the fit. Section 4.5 concludes with a brief summary and some suggestions for future research.

4.1 Modeling Returns, Bipower Variation and Jumps

A thorough discussion of the most popular realized volatility models has been provided in Section 3.1. While these models describe the dynamics of the total price variation, we explicitly decompose here the volatility into its two components, and formulate a model for returns, Bipower variation and the relative jump series. However, as the Bipower variation and the realized volatility exhibit similar characteristics (see Section 2.2.2) we can adopt the approaches developed for realized volatility models to the Bipower variation. We therefore start the discussion of our multivariate model with the specification of the Bipower variation equation, proceed with the jump equation and then introduce the return equation.

4.1.1 The BV Equation

From the previous discussions we know, that Bipower variation is a component of realized volatility and that the two only differ (by more than measurement error) in the presence of jumps. Moreover, we have seen that both series exhibit strong persistence and are approximately lognormally distributed. As such, an ARFIMA or HAR specification may be adequate for modeling logarithmic Bipower variation. Given their equally well performance in modeling and forecasting realized volatility, we follow the approach adopted in Section 3.2 and specify a HAR specification. This choice is primarily driven by its simplicity in estimation. Based on our findings from the previous chapter, we also allow for time–variation in the volatility of the continuous volatility component and include a GARCH specification.

To set up the model we define the logarithmic multiperiod Bipower variation measures by the sum of the corresponding daily logarithmic measures

$$(\log BV)_{t+1-k:t} = \frac{1}{k}\sum_{j=1}^{k} \log BV_{t-j}, \tag{4.1}$$

where $k = 5$ and $k = 22$ correspond to (approximately) one week and one month, respectively.[2] Our HAR–GARCH–BV model can then be formulated as

$$\begin{aligned} \log BV_t &= \alpha_0 + \alpha_d \log BV_{t-1} + \alpha_w (\log BV)_{t-5:t-1} + \alpha_m (\log BV)_{t-22:t-1} \\ &\quad + \theta_1 \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} + \theta_2 I[r_{t-1} < 0] + \theta_3 \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} I[r_{t-1} < 0] + \sqrt{h_t} u_t \end{aligned} \tag{4.2}$$

$$\begin{aligned} h_t &= \omega + \sum_{j=1}^{q} \alpha_j (\log BV_{t-1} - x'_{BV,t-1}\beta_{BV})^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \\ &\quad + \sum_{j=1}^{s} \lambda_j BV_{t-j}. \end{aligned} \tag{4.3}$$

[2] We follow Corsi (2004) in defining the multi–period logarithmic volatility by the sum of the corresponding one–period logarithmic measures. Almost identical empirical results are obtained by using the logarithm of the multi–period realized variances in place of the sum of the logarithms.

The lagged daily, weekly and monthly realized variation measures on the right–hand–side of the $\log BV_t$ equation could, of course, be augmented with additional terms to account for the possibility of even longer–run dependencies. However, the combination of relatively few volatility components often provide a remarkably close approximation to true long–memory dependencies. Equation (4.2) still differs slightly from the HAR–GARCH specification for the realized volatility given in equation (3.4). In particular, we have additionally included lagged signed returns, which should account for a leverage effect in the continuous volatility component. Moreover, motivated by the observation in Barndorff-Nielsen and Shephard (2005) that the volatility of realized volatility tends to be high when the volatility is high, the model also permits a level effect in the GARCH model for the volatility of the continuous volatility component.

Lastly, the descriptive analysis in Section 2.2.2 suggests that Bipower variation is only approximately lognormally distributed. To account for such deviations we assume a normal–mixture distribution for the innovations:

$$u_t \overset{i.i.d.}{\sim} \begin{cases} \mathbb{N}_1(0,1) & \text{with probability } (1-p_{u,2}) \\ \mathbb{N}_2(\mu_{u,2}, \sigma^2_{u,2}) & \text{with probability } p_{u,2} \end{cases} . \tag{4.4}$$

Note, that in our empirical application we have also experimented with the use of a standard normal distribution. This distribution, however, was decisively rejected against the normal–mixture distribution by a standard Likelihood–Ratio test. Having defined the process of the continuous volatility component we now turn to the specification of the jump part.

4.1.2 The Jump Equation

As highlighted in the descriptive analysis in Section 2.2.2, the jumps exhibit weak, but non–zero, autocorrelation. We account for this by specifying a standard autoregressive model

$$\begin{aligned} \log\left(\frac{RV_t}{BV_t}\right) &= \delta_0 + \psi_1 \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} + \psi_2 I[r_{t-1}<0] + \psi_3 \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} I[r_{t-1}<0] \\ &\quad + \sum_{j=1}^{n} \delta_j \log\left(\frac{RV_{t-j}}{BV_{t-j}}\right) + \nu_t, \end{aligned} \tag{4.5}$$

where we additionally include the same leverage specification as in the Bipower variation equation. The latter is an interesting point, as it allows us to disentangle whether the often observed asymmetric negative relationship between total volatility and return innovations is primarily driven by the response of the continuous volatility component and/or by the reaction of jumps to the arrival of news.

The innovations in the jump equation are assumed to follow a mixture of a zero mean Normal Inverse Gaussian (NIG) distribution and an Inverse Gaussian (IG) distribution:

$$\nu_t \overset{i.i.d.}{\sim} \begin{cases} \mathbb{NIG}_0(\alpha_{NIG}, \beta_{NIG}, \delta_{NIG}) & \text{with probability } (1-p_{\nu,2}) \\ \mathbb{IG}(\lambda_{IG}, \mu_{IG}) & \text{with probability } p_{\nu,2} \end{cases} . \tag{4.6}$$

Obviously, the jump innovations would be misspecified by a Gaussian distribution due to the prevalent large skewness and fat right tail behavior in the unconditional distribution of the jump measure. Other distributions, such as the skewed student–t distribution, could of course be considered as well. However, experimentation suggests that the above mixture specification provides an accurate description of the jump innovations. Moreover, a mixture of distributions with one distribution having support over the whole real line and the other being defined only on the positive domain seems to be a sensible approach in that one distribution is primarily capturing the small fluctuations of logarithmic realized variance around logarithmic Bipower variation, which are attributable to measurements error and small jumps, whereas the second distribution captures the large jumps, i.e. the relevant right part of the distribution. Intuitively, the mixing probability $p_{\nu,2}$ then gives an indication of the relevance of large genuine jumps. Furthermore, our particular choice of the NIG and IG distributions is motivated by their ability to produce quite flexible shapes. The accuracy of this particular distribution of mixtures is documented in Section 4.2.

4.1.3 The Return Equation

Our return equation is quite standard. In particular, we specify an autoregressive process to account for some short–term dynamics in the daily returns. However, the innovations are scaled by the total price variation as measured by the realized volatility. Although the latter is not modeled directly, it can straightforwardly be inferred from the Bipower variation and jump equations using the definition of the relative jump measure, i.e. $RV_t := \exp(J_t + \log BV_t)$. Hence, the return process is given by

$$r_t = \gamma_0 + \sum_{j=1}^{d} \gamma_j r_{t-j} + \sqrt{RV_t}\epsilon_t, \tag{4.7}$$

where we assume standard normally distributed innovations

$$\epsilon_t \overset{i.i.d.}{\sim} \mathbb{N}(0,1). \tag{4.8}$$

Note that from the theoretical point of view the distribution of the return innovations may differ from normality, depending on the relevance of the jumps and their respective distribution. We have therefore also experimented with more flexible distributions, such as the NIG or a mixture of normals. However, the standard normal distribution provides a similarly accurate fit, which is also consistent with the descriptive statistics presented in Section 2.2.2. The next section presents the estimation results of the three equations.

4.2 Equation–by–Equation Estimation

The recursive structure of the three equation system defined in the preceding section, means that as long as the disturbances are independent across equations, each of the three models may be estimated efficiently in isolation using standard maximum–likelihood methods. The assumption of independent disturbances is, of course, questionable, as the stochastic volatility literature points towards the existence of contemporaneous dependencies such as a leverage or volatility feedback effect. We therefore explicitly investigate the validity of this assumption based upon the single equation estimates. This approach also allows us to assess the form of such effects more closely and to disentangle their importance for the two volatility components. The next section presents the estimation results, while Section 4.2.2 analyzes the residual dependencies.

4.2.1 Equation–by–Equation Estimation Results

The parameter estimates for each of the three equations, along with the corresponding asymptotic standard errors, are reported in Table 4.1.[3] Figures 4.1 to 4.3 show the resulting residuals, their autocorrelation and partial autocorrelation functions, as well as the QQ plots and kernel density estimates. The selection of the autoregressive lags in the different models is based on the Schwarz Bayesian information Criterion (BIC), and all of the lags are kept the same in the subsequent models.

Starting with the results in the first column and the BV_t equation, the estimates directly mirror earlier results in the literature for the HAR realized volatility model. The daily, weekly and monthly volatility components are all highly statistically significant, while the inclusion of the logarithmic Bipower variation measures over biweekly and other horizons do not improve the fit according to the BIC criteria. A standard GARCH(1,1) model without any level effects emerges as the preferred specification for the conditional variance.[4] The estimated GARCH parameters easily satisfy the corresponding stationarity condition $\alpha_1\sigma_u^2 + \beta_1 < 1$, where $\sigma_u^2 = 1 + p_{u,2}\left(\sigma_{u,2}^2 - 1\right)$. Figure 4.4 depicts the time evolvement of the filtered volatility of the Bipower variation showing that the volatility of the continuous volatility component exhibits similarl but less pronounced time–varying patterns as found for the volatility of realized volatility (see Chapter 3). The asymmetry, or leverage effect, in the continuous volatility component is directly manifest by the highly significant estimates for the θ_1 and θ_3 parameters. As expected, the point estimates

[3] Note that, although mixtures of distributions can sometimes be difficult to estimate, we did not encounter any convergence problems. Also, to ensure proper convergence we estimated each of the equations based upon a range of different starting values.

[4] According to the BIC a GARCH(1,1) is sufficient although the ARCH–LM test and PACF of the squared innovations weakly indicated a GARCH(1,2) specification. The inclusion of a second ARCH component, however, is insignificant. The same holds true for the level effect in the GARCH equation and we have therefore excluded it. From the estimation results we also find that the lagged conditional variance is more important than the squared Bipower variation innovations.

Table 4.1: Single Equation Estimation Results

	BV equation			Jump equation			Return equation	
	Estimate	Std. Error		Estimate	Std. Error		Estimate	Std. Error
α_0	-0.1978	(0.0170)	δ_0	0.0704	(0.0067)	γ_0	0.0858	(0.0098)
α_d	0.2548	(0.0169)	δ_1	0.0347	(0.0089)	γ_2	-0.0254	(0.0139)
α_w	0.4370	(0.0265)	δ_5	0.0516	(0.0116)	γ_3	-0.0351	(0.0133)
α_m	0.2416	(0.0215)	ψ_1	-0.0143	(0.0032)			
θ_1	0.0571	(0.0144)	ψ_2	-0.0026	(0.0050)			
θ_2	0.0384	(0.0217)	ψ_3	0.0014	(0.0049)			
θ_3	0.1247	(0.0218)	$p_{\nu,2}$	0.0072	(0.0329)			
ω	0.0228	(0.0053)	α_{NIG}	71.5659	(52.7253)			
α_1	0.0419	(0.0077)	β_{NIG}	54.0383	(47.7732)			
β_1	0.8048	(0.0378)	δ_{NIG}	0.2637	(0.0367)			
$p_{u,2}$	0.1451	(0.0304)	λ_{IG}	0.5247	(0.3198)			
$\mu_{u,2}$	0.7688	(0.1306)	μ_{IG}	1.1804	(5.2968)			
$\sigma_{u,2}$	1.9278	(0.0688)						
logL:	-3464.75		logL:	3775.22		logL:	-5839.63	

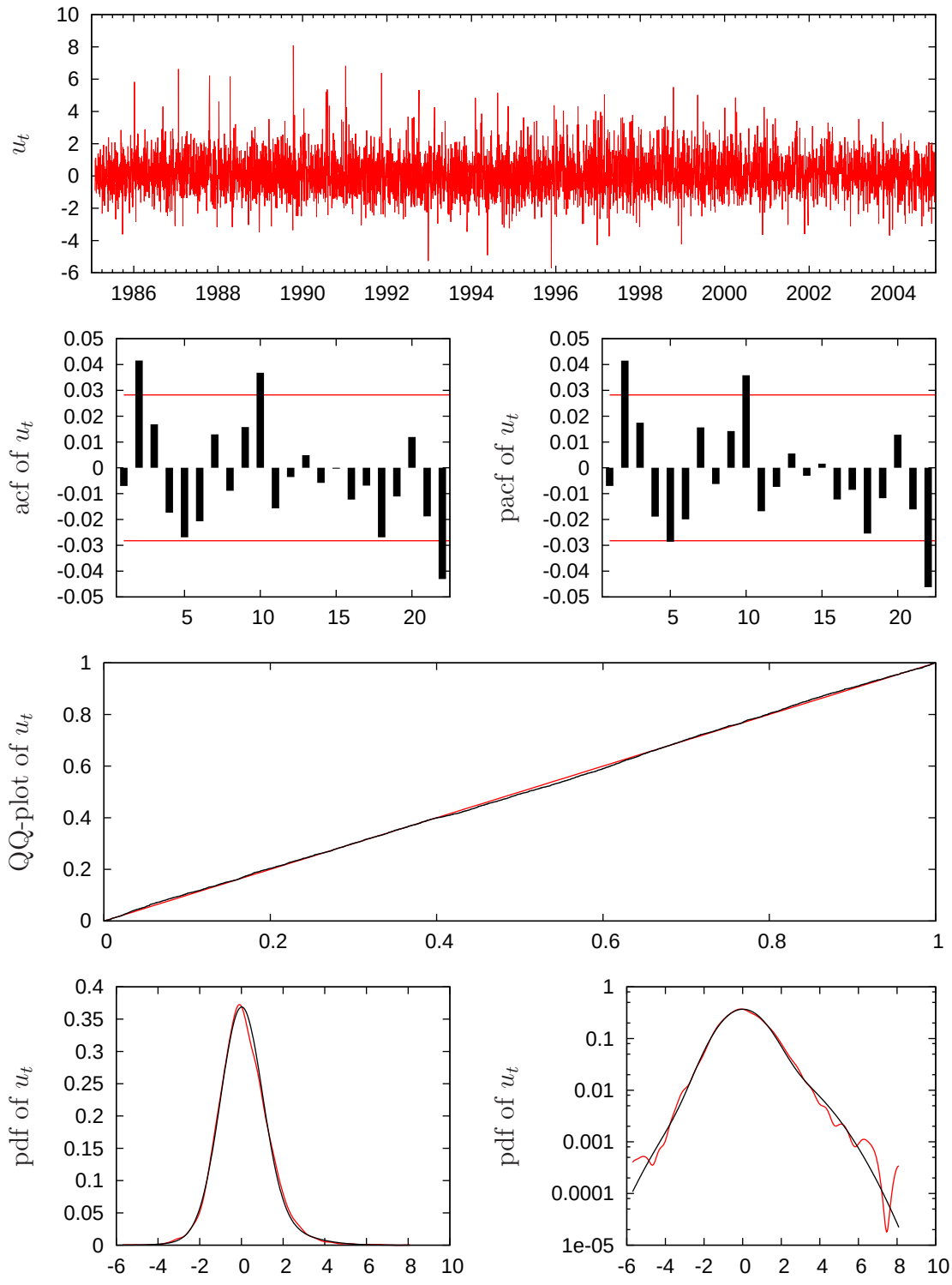

Figure 4.1: Residual analysis of the (log.) Bipower variation equation. The upper graph of the figure represents the time evolvement of the innovations of the Bipower variation equation. The second line of graphs shows their sample autocorrelations and partial autocorrelations. The third is the corresponding Quantile–Quantile plot. The lower left panel of the figure shows the kernel density estimates of the residuals (dashed line) and the density of the estimated normal mixture (solid line). The right panel shows the same in log scale.

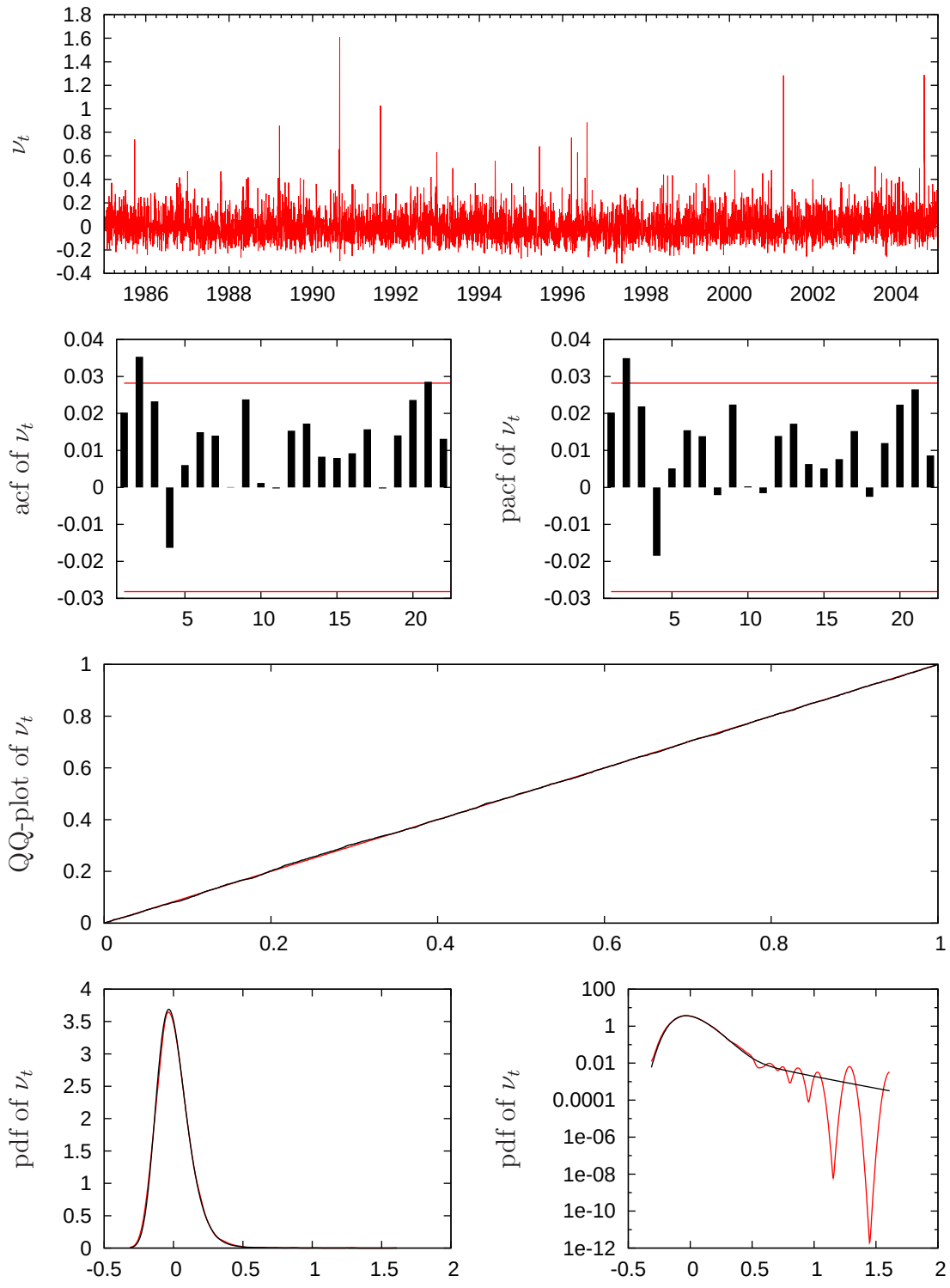

Figure 4.2: Residual analysis of the jump equation. The upper graph of the figure represents the time evolvement of the innovations of the jump equation. The second line of graphs shows their sample autocorrelations and partial autocorrelations. The third is the corresponding Quantile–Quantile plot. The lower left panel of the figure shows the kernel density estimates of the residuals (dashed line) and the density of the estimated NIG–IG mixture (solid line). The right panel shows the same in log scale.

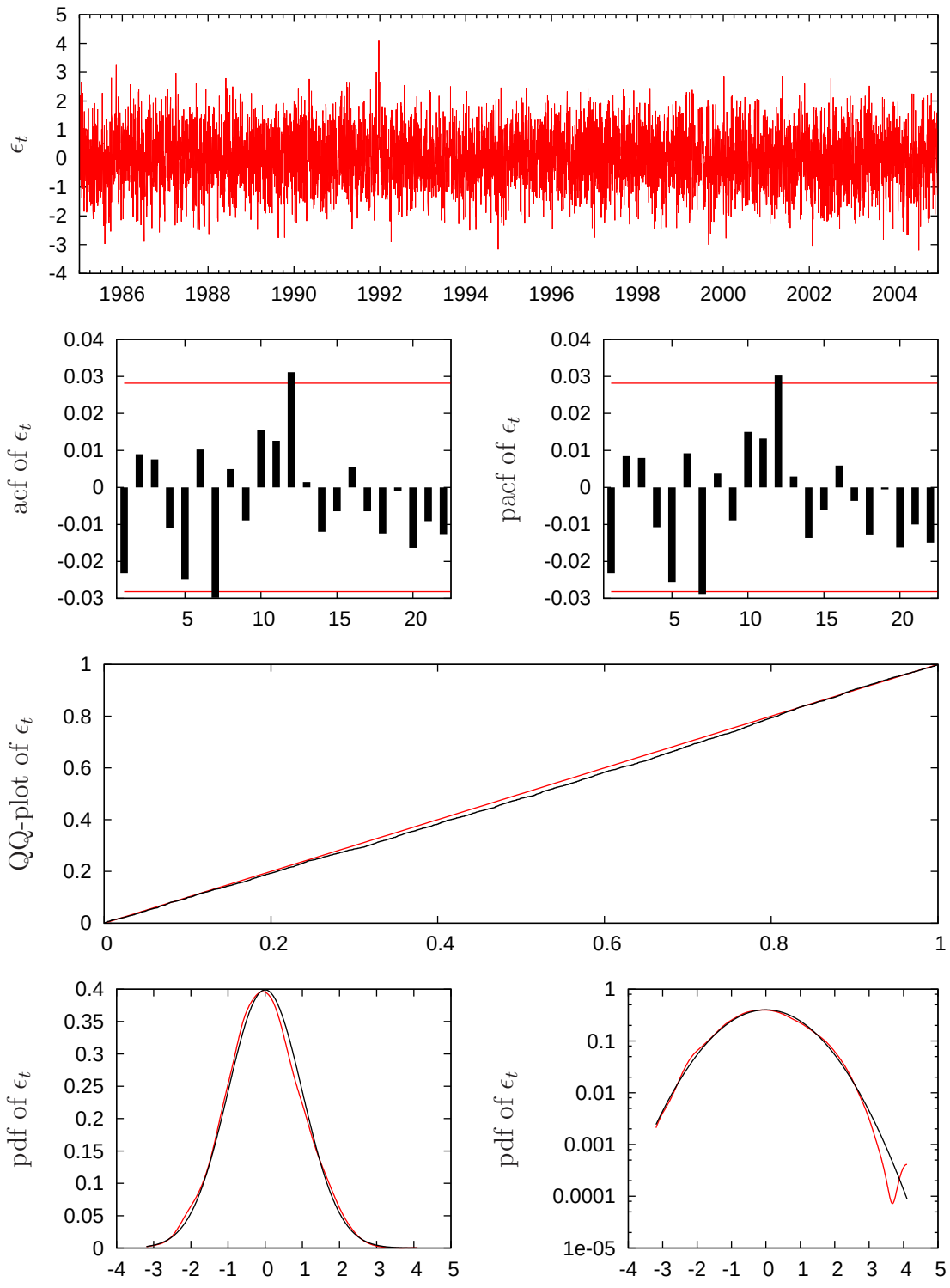

Figure 4.3: Residual analysis of the return equation. The upper graph of the figure represents the time evolvement of the innovations of the return equation. The second line of graphs shows their sample autocorrelations and partial autocorrelations. The third is the corresponding Quantile–Quantile plot. The lower left panel of the figure shows the kernel density estimates of the residuals (dashed line) and the density of a standard normal (solid line). The right panel shows the same in log scale.

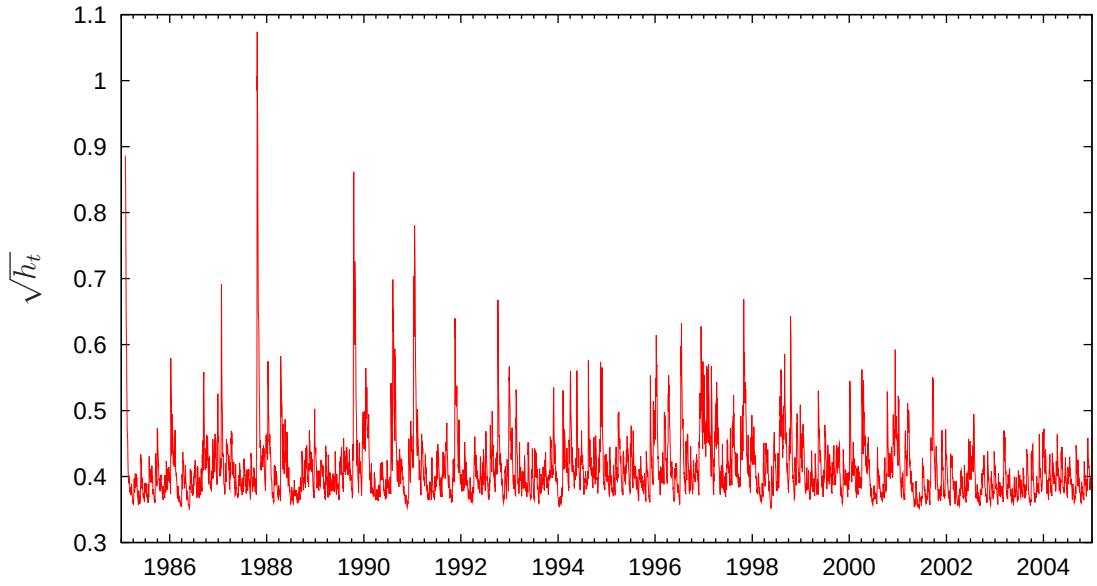

Figure 4.4: The volatility of Bipower variation. The graph exhibits the HAR–GARCH implied volatility series of logarithmic Bipower variation.

imply that a lagged negative return shock leads to a much larger increase in the volatility than does a positive shock of the same magnitude. In contrast, the level shift in the volatility equation due to negative news is not significant. This latter result is consistent with earlier findings for the realized volatility in Martens et al. (2004). The QQ and kernel density plots in Figure 4.1 also indicate that the mixture of two normal distributions does a very good job of capturing the slight skewness and kurtosis inherent in the innovations from the model. Moreover, the autocorrelation and partial autocorrelation functions for the estimated residuals do not reveal any remaining systematic serial correlation within a monthly horizon.

Turning to the jump equation, the autoregressive parameter estimates associated with the first, or daily, and fifth, or weekly, lags are both significant. Still, the magnitude of both coefficients is very small, thus supporting the aforementioned weak own predictability in the jump series. Interestingly, and in sharp contrast to the results for the continuous volatility component, the parameter estimates for ψ_2 and ψ_3 related to the leverage effect suggest that jumps are *not* asymmetrically affected by lagged return shocks. These results also bear important implications, i.e. we find evidence that a negative leverage effect is only present in the diffusive volatility component but not in the jump part. In fact, this supports most of the continuous–time jump diffusion models, in which the leverage effect is usually imposed by a negative correlation between the two Brownian motions driving the price and volatility processes, see e.g. the models in Bates (2000), Eraker et al. (2003) and Pan (2002). Our results on the contemporaneous dependencies in the distur-

bances presented in the next section point into the same direction. Moreover, note that if anything the estimate for ψ_1 points to a symmetric, but dampening impact of news on future jumps.[5] The QQ–Plot for the residuals from the J_t equation as well as the kernel density plots in Figure 4.2 also show that the distribution of the jump innovations is well described by the NIG–IG mixture. Based on the very low mixing probability of the IG (i.e. of $p_{nu,2}$) one might conjecture, that the mixing is not relevant and a single NIG distribution could be sufficient. Although we believe that a mixture is required to reproduce such a density shape, we have also tried the pure NIG distribution and had to reject it based on the Likelihood–Ratio test comparing the NIG with the NIG–IG specification. The BIC and the density plots also point to the adequacy of the NIG–IG model.

The estimates for the return equation, reported in the last column, reveal statistically significant, but economically very small, second and third order autocorrelations. As already noted, the standard normal distribution appears to fit the data well, and it is generally preferred over other specifications by the BIC criteria, including a normal distribution with a freely estimated variance as well as a freely estimated zero–mean NIG distribution. We also experimented with the inclusion of a risk premia, or GARCH–in–Mean type effect, by allowing the conditional mean to depend on the realized variance. Consistent with existing results in the literature suggesting that reliable estimates for this risk premium parameter requires longer return horizons and time-spans of data (see e.g. Lundblad (2004) and Ghysels et al. (2005)), we found the GARCH–in–Mean effect to be insignificant at the daily level.

4.2.2 Residual Inter–Dependencies

The separate estimation of the three equations discussed above implicitly assumes that the disturbances are independent. However, based upon existing results in the stochastic volatility literature, we might naturally expect that the disturbances in the return and volatility equations are correlated due to contemporaneous (at the daily level) leverage and/or volatility feedback effects; see, e.g., the recent empirical analysis in Bollerslev et al. (2006b). Moreover, the innovations to the two volatility equations might naturally be expected to be correlated as well. Such inter–dependencies would obviously have to be taken into account in a fully efficient estimation of the joint system, and could in principle result in inconsistent equation–by–equation estimates.

To begin, consider the sample correlation matrix for the estimated residuals from

[5] In the context of a representative agent general equilibrium model, Tauchen (2005) has recently shown that a positive leverage effect can occur depending on the magnitude of the intertemporal marginal rate of substitution and the degree of risk aversion. It is possible that by explicitly differentiating between the two sources of risk, an extension of this model could help explain our empirical findings of a "standard" negative leverage effect in the diffusion component but a positive correlation between returns and jumps.

the Bipower variation, jump and return equations

$$\hat{\rho} = \begin{bmatrix} 1 & -0.1847 & -0.2008 \\ . & 1 & 0.0283 \\ . & . & 1 \end{bmatrix}.$$

Consistent with the discussion above, the continuous volatility innovations appear to be negatively correlated with both the relative jump residuals and the return innovations. Meanwhile, the correlation between the relative jumps and the return residuals appears negligible.

In addition to the linear contemporaneous relationships suggested by the sample correlations, there might also exist non–linear dependencies due to, e.g., asymmetric volatility effects. Figures 4.5 to 4.7 present the pairwise scatter plots of the residual series along with a fitted quadratic polynomial, as well as a Rosenblatt–Parzen Gaussian–based kernel estimator. The conjecture of a nonlinear relationship between the residuals is seemingly evident for at least two of the three combinations. Most obviously, there is an asymmetric negative relation between the residuals of the Bipower equation and the return shocks in Figure 4.5.[6] In fact, this relationship is very similar to the commonly assumed lagged leverage effect. In contrast, there is no apparent non–linear relation between the residuals from the jump and return equations. Interestingly, Figure 4.7 reveals a smirk–like relation between the innovations to the continuous volatility and jump components. This effect should, of course, be carefully interpreted in light of the definitions of the underlying variation measures. In particular, a negative shock to the (logarithmic) Bipower variation corresponds to an overestimation of the continuous volatility component, which in turn is associated with a larger jump component. In contrast, a positive shock to the Bipower variation equation, and a larger than expected continuous volatility component, does not directly affect the relative jump measure.

To further visualize the inter-dependencies between the estimated residuals, Figure 4.8 shows the scatter plot of the respective pairwise probability integral transform, or PIT, series defined as the cumulative distribution function (cdf) evaluated at the realized innovations.[7] In the absence of any inter–dependencies and for correctly specified marginal innovation densities, the points should be uniformly distributed over the whole scatter surface. Consistent with the aforementioned smile–like pattern in the residual scatter plot for the Bipower variation and return equations, the first panel shows that low (high) cdf values of the return innovations tend to be associated with higher (medium) cdf values of the innovation to the continuous volatility component. A similar pattern emerges in the cdf scatter for

[6] The estimated parameters of the fitted quadratic polynomials, with corresponding HAC robust standard errors in parenthesis, equal -0.0199 (0.0210), -0.2430 (0.0179) and 0.1288 (0.0143), respectively, being indicative of a highly statistically significant asymmetric relationship.

[7] Recall, that within the density forecasting context, we have already provided a formal definition of the probability integral transform. Equation (3.6) also holds here, but y_t now denotes the realized innovation and $f(t)$ is the assumed distribution of the innovations being independent of the past.

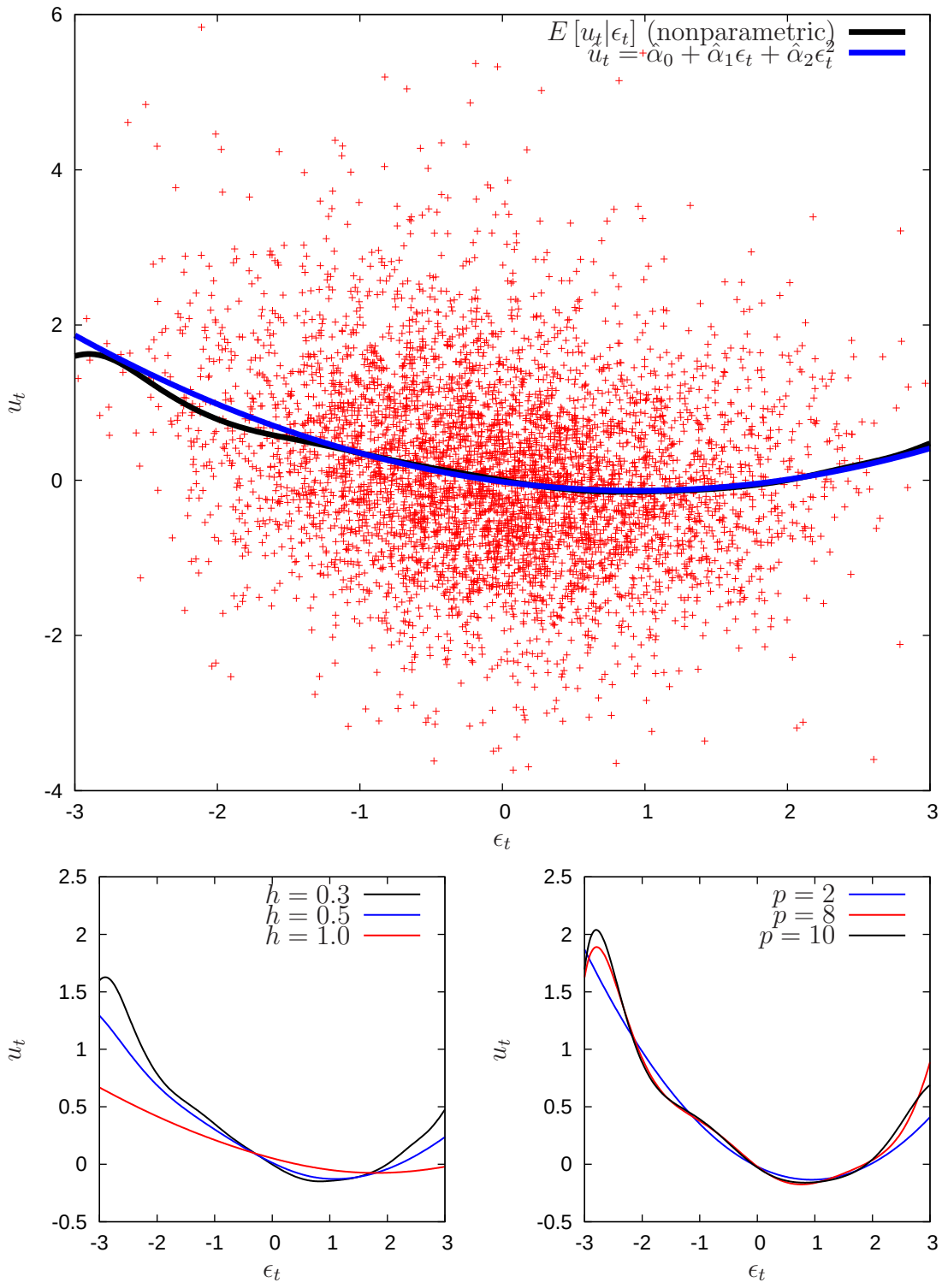

Figure 4.5: Dependency analysis of the residuals between the return equation and the Bipower variation equation. The lower left and right panels include additional different polynomial and non–parametric specifications, respectively.

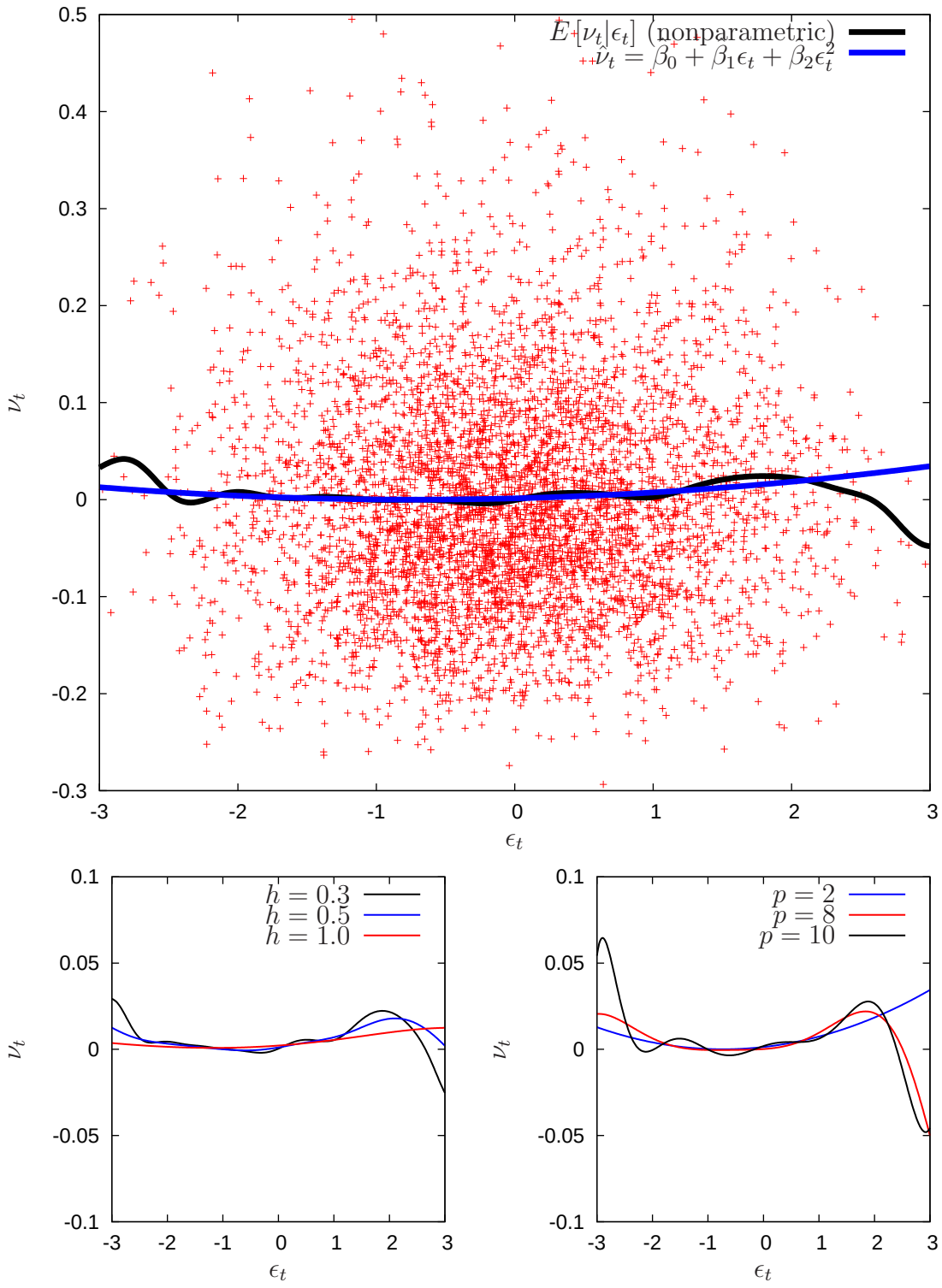

Figure 4.6: Dependency analysis of the residuals between the return equation and the jump equation. The lower left and right panels include additional different polynomial and non–parametric specifications, respectively.

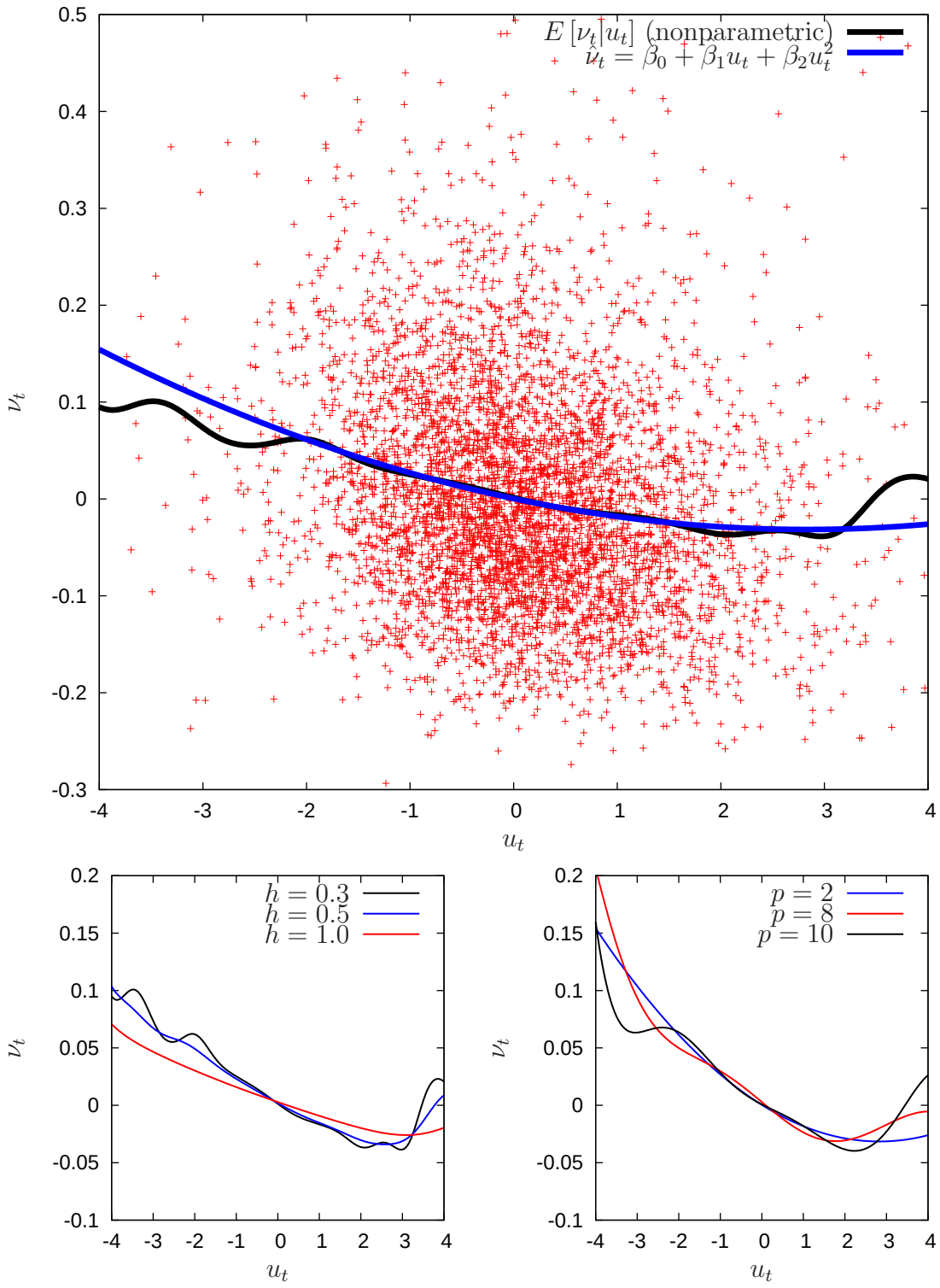

Figure 4.7: Dependency analysis of the residuals between the Bipower variation equation and the jump equation. The lower left and right panels include additional different polynomial and non–parametric specifications, respectively.

the jump and continuous volatility innovations in the bottom panel, but with high return cdf values being associated with smaller values of the jump innovation cdf due to the dampening (smirk–like) behavior. Meanwhile, the cdf scatter between the jump and return innovations in the middle panel exhibits nearly uniformly distributed scatter points.

In summary, our analysis points to the existence of important asymmetric dependencies among the three innovation series. These effects should be incorporated into a joint modeling framework in order to, firstly, more systematically quantify and test for their significance, secondly, guard against any biases in the single equation estimates, and thirdly, enhance the efficiency of the individual model parameter estimates. The unified system approach explicitly allowing for non–linear functional forms of residual dependencies presented in the next section accomplishes these goals.

4.3 System Estimation

The results of the equation–by–equation estimations suggest that the proposed model specifications provide an adequate description of the dynamic dependencies in the two volatility and return processes, but that it does not fully account for the nonlinear contemporaneous dependencies among the innovations. We therefore retain our basic three equation set up, but additionally model the nonlinear interdependencies based on the following system of equations

$$
\begin{aligned}
r_t &= \gamma_0 + \sum_{j=1}^{d} \gamma_j r_{t-j} + \sqrt{RV_t}\epsilon_t \\
\log BV_t &= \alpha_0 + \alpha_d \log BV_{t-1} + \alpha_w (\log BV)_{t-5:t-1} + \alpha_m (\log BV)_{t-22:t-1} \\
&\quad + \theta_1 \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} + \theta_2 I[r_{t-1}<0] + \theta_3 \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} I[r_{t-1}<0] + \sqrt{h_t}\,(u_t + g(\epsilon_t)) \\
h_t &= \omega + \sum_{j=1}^{q} \alpha_j \left(\log BV_{t-j} - x'_{BV}\beta_{BV}\right)^2 + \sum_{j=1}^{p} \beta_j h_{t-j} + \sum_{j=1}^{s} \lambda_j BV_{t-j} \\
\log\left(\frac{RV_t}{BV_t}\right) &= \delta_0 + \sum_{j=1}^{n} \delta_j \log\left(\frac{RV_{t-j}}{BV_{t-j}}\right) \\
&\quad + \psi_1 \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} + \psi_2 I[r_{t-1}<0] + \psi_3 \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} I[r_{t-1}<0] \\
&\quad + \left(\nu_t + m(u_t) + k(\epsilon_t)\right).
\end{aligned}
\tag{4.9}
$$

In comparison to the individual equations, the system explicitly allows the innovations in the continuous volatility and relative jump equations to depend nonlinearly on the return innovations via the general functions $g(\epsilon_t)$ and $k(\epsilon_t)$, respectively. Similarly, the jump innovations are allowed to depend on the continuous volatility

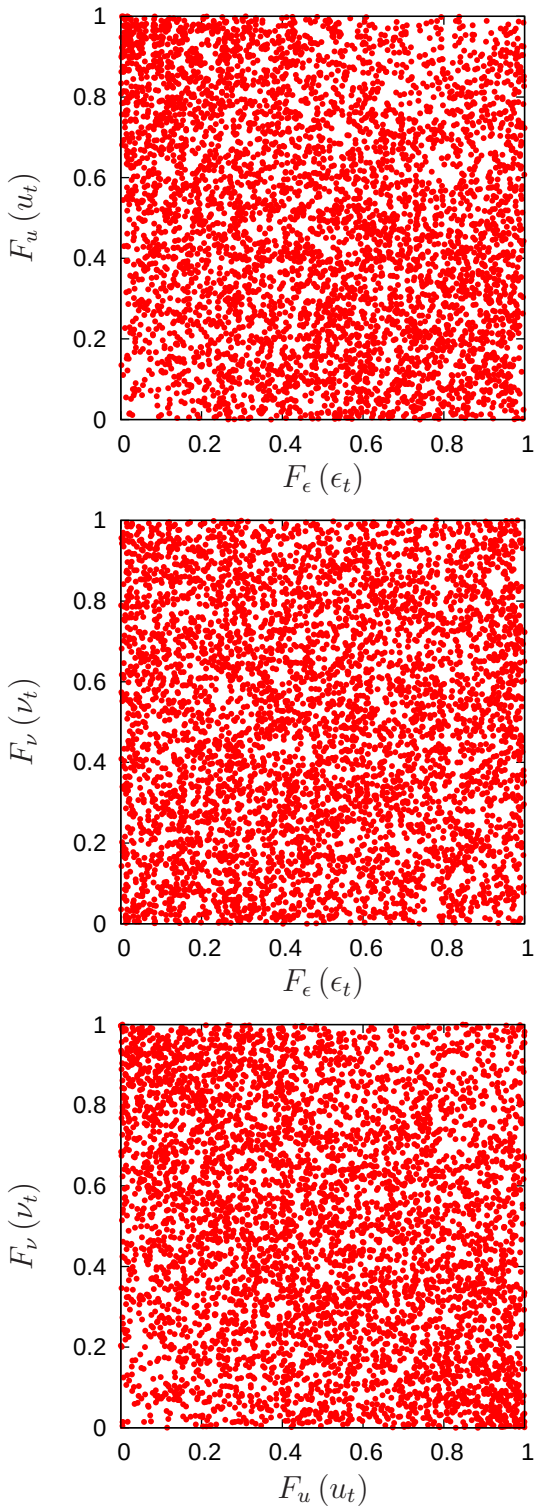

Figure 4.8: Realized CDF scatter plot of the single equation innovations.

shocks via the $m(u_t)$ function. Thus, by choosing an adequate functional form for each of these functions, we seek to render the underlying three innovation series to be pairwise independent.

Given the density functions of the disturbances, we can derive the transition density of the system. Note that the recursive structure of both the basic model equations as well as of the inter–dependencies strongly simplify the computations. So, utilizing the contemporaneous independence of the transformed innovations, the transition density for the joint system, $y_t = (\log BV_t, \log\left(\frac{RV_t}{BV_t}\right), r_t)'$, may be readily expressed as

$$
\begin{aligned}
f_y(y_t|x_{t-1};\theta) = & \frac{1}{\sqrt{h_t}\sqrt{RV_t}} \\
& \times f_\epsilon\left(\underbrace{\frac{r_t - x_r'\beta_r}{\sqrt{RV_t}}}_{\epsilon_t}\middle|\,\vartheta_\epsilon\right) f_u\left(\underbrace{\frac{\log BV_t - x_{BV}'\beta_{BV}}{\sqrt{h_t}} - g\left(\frac{r_t - x_r'\beta_r}{\exp\left\{\frac{1}{2}\log RV_t\right\}}\right)}_{u_t}\middle|\,\vartheta_u\right) \\
& \times f_\nu\left(\underbrace{\log\left(\frac{RV_t}{BV_t}\right) - x_{RV}'\beta_{RV} - m(u_t) - k(\epsilon_t)}_{\nu_t}\middle|\,\vartheta_\nu\right),
\end{aligned}
$$

where as before

$$
\begin{aligned}
\epsilon_t &\overset{i.i.d.}{\sim} \mathbb{N}(0,1) \\
u_t &\overset{i.i.d.}{\sim} \begin{cases} \mathbb{N}_1(0,1) & \text{with probability } (1-p_{u,2}) \\ \mathbb{N}_2(\mu_{u,2}, \sigma_{u,2}^2) & \text{with probability } p_{u,2} \end{cases} \\
\nu_t &\overset{i.i.d.}{\sim} \begin{cases} \mathbb{NIG}_0(\alpha_{NIG}, \beta_{NIG}, \delta_{NIG}) & \text{with probability } (1-p_{\nu,2}) \\ \mathbb{IG}(\lambda_{IG}, \mu_{IG}) & \text{with probability } p_{\nu,2}. \end{cases}
\end{aligned}
$$

To complete the specification, we assume that the nonlinear contemporaneous dependencies among the individual equation innovations may be adequately captured by a set of second order degree polynomials[8]

$$
g(\epsilon_t) = g_1\epsilon_t + g_2\epsilon_t^2 \tag{4.10}
$$

$$
k(\epsilon_t) = k_1\epsilon_t + k_2\epsilon_t^2 \tag{4.11}
$$

$$
m(u_t) = m_1 u_t + m_2 u_t^2, \tag{4.12}
$$

[8]We also experimented with higher order polynomials, but found the simple quadratic representations to be sufficient in capturing the smirk–like dependencies over the required range.

where for identification purposes we have restricted the three constants to be zero. Fully efficient maximum likelihood estimation of the complete system may now proceed in a standard manner by maximizing the log likelihood function defined by the summation of the logarithmic transition densities over the sample observations.[9]

Comparing the system estimation results reported in Table 4.2 to the equation–by–equation results in Table 4.1, the estimates for most of the individual parameters obviously do not change by much.[10] In particular, our previous conclusions regarding the lagged leverage effect in the continuous volatility component and the positive correlation between jump and return innovations all remain intact.[11] Moreover, as expected the asymptotic standard errors for the estimated parameters are generally smaller for the system estimates in Table 4.2, highlighting the gain in (asymptotic) efficiency obtained by jointly estimating the three equations.

Our major interest, however, concerns the estimation results of the dependency functions. In fact the parameter estimates support our earlier findings based on the visual inspection of the single equation estimation. The highly significant quadratic term in the $g(\epsilon_t)$ dependency function clearly indicates that the innovations to the continuous volatility component are non–linearly related to the innovations to the return equation. In contrast, such dependence cannot be found for the jump component, for which only k_1 is significant, but numerically very small. More specifically, excluding the non–linear coefficient, i.e. $k_2 = 0$, in the system yields slightly significant improvements in the model as measured by the BIC criterium and the Likelihood-Ratio test. The aforementioned non–linear relationship between the continuous volatility component and the relative jump innovations allowed for by the $m(u_t)$ dependency function is also strongly supported by the joint estimation. The importance of allowing for contemporaneous non–linear dependencies among the innovations is further underscored by the Likelihood–Ratio test comparing the fully specified simultaneous equation model to the system equation estimates without the quadratic polynomials, which equals an overwhelmingly significant 597.67.

The model presented in Table 4.2 still includes some individually insignificant parameters. In particular, restricting $\theta_2 = \psi_2 = \psi_3 = k_2 = 0$, and re–estimating the model results in a Likelihood–Ratio test statistic of only 6.619 versus the fully general model. Also, the remaining parameter estimates are hardly affected by restricting these four parameters to equal zero. Our final preferred model specification is therefore given by this restricted model as presented in Table 4.3.

Our estimation results, however, are based on the assumption of independently

[9] The model was initially estimated using the estimates from the last section as starting values. But to ensure proper convergence we also estimated the model with a series of different starting values, resulting in identical numerical values.

[10] Further analysis related to the dynamic dependencies and unconditional distributional properties of the system residuals also yield almost identical results to the ones for the single–equation residuals in Figures 4.1 to 4.3.

[11] Importantly, the system GARCH parameter estimates for the BV_t equations also satisfy the corresponding second-order stationarity condition: $\alpha\left(\sigma_u^2 + g_1^2 + 2g_2^2\right) + \beta < 1$, where $\sigma_u^2 = 1 + p_{u,2}\left(\sigma_{u,2}^2 - 1\right)$ and $\alpha \geq 0$, $\beta \geq 0$.

Table 4.2: System Estimation Results (logL=-5230.37)

	BV equation			Jump equation			Return equation	
	Estimate	Std. Error		Estimate	Std. Error		Estimate	Std. Error
α_0	-0.2526	(0.0172)	δ_0	0.0665	(0.0051)	γ_0	0.0570	(0.0095)
α_d	0.2499	(0.0160)	δ_1	0.0422	(0.0095)	γ_2	-0.0321	(0.0125)
α_w	0.4494	(0.0249)	δ_5	0.0500	(0.0110)	γ_3	-0.0431	(0.0116)
α_m	0.2291	(0.0205)	ψ_1	-0.0145	(0.0033)			
θ_1	0.0636	(0.0139)	ψ_2	-0.0034	(0.0050)			
θ_2	0.0424	(0.0215)	ψ_3	0.0028	(0.0051)			
θ_3	0.1246	(0.0211)	m_1	-0.0200	(0.0012)			
g_1	-0.2493	(0.0186)	m_2	0.0013	(0.0004)			
g_2	0.1363	(0.0129)	k_1	0.0042	(0.0015)			
ω	0.0250	(0.0055)	k_2	0.0018	(0.0011)			
α_1	0.0425	(0.0077)	$p_{\nu,2}$	0.0174	(0.0263)			
β_1	0.7707	(0.0417)	α_{NIG}	41.8149	(13.6452)			
$p_{u,2}$	0.1617	(0.0035)	β_{NIG}	26.1884	(11.2286)			
$\mu_{u,2}$	0.6183	(0.1204)	δ_{NIG}	0.2417	(0.0330)			
$\sigma_{u,2}$	1.9391	(0.0731)	λ_{IG}	0.3183	(0.0933)			
			μ_{IG}	0.3722	(0.7001)			

Table 4.3: Restricted System Estimation Results (logL=-5233.67)

	BV equation			Jump equation			Return equation	
	Estimate	Std. Error		Estimate	Std. Error		Estimate	Std. Error
α_0	-0.2351	(0.0140)	δ_0	0.0668	(0.0042)	γ_0	0.0572	(0.0095)
α_d	0.2510	(0.0160)	δ_1	0.0426	(0.0094)	γ_2	-0.0323	(0.0125)
α_w	0.4476	(0.0249)	δ_5	0.0497	(0.0110)	γ_3	-0.0430	(0.0116)
α_m	0.2298	(0.0205)	ψ_1	-0.0136	(0.0025)			
θ_1	0.0489	(0.0115)	ψ_2	-	-			
θ_2	-	-	ψ_3	-	-			
θ_3	0.1596	(0.0126)	m_1	-0.0200	(0.0012)			
g_1	-0.2493	(0.0186)	m_2	0.0013	(0.0004)			
g_2	0.1406	(0.0127)	k_1	0.0045	(0.0015)			
ω	0.0247	(0.0055)	k_2	-	-			
α_1	0.0419	(0.0077)	$p_{\nu,2}$	0.0198	(0.0236)			
β_1	0.7728	(0.0416)	α_{NIG}	41.0467	(12.6795)			
$p_{u,2}$	0.1628	(0.0355)	β_{NIG}	25.6054	(10.3213)			
$\mu_{u,2}$	0.6149	(0.1194)	δ_{NIG}	0.2390	(0.0322)			
$\sigma_{u,2}$	1.9374	(0.0730)	λ_{IG}	0.3007	(0.0830)			
			μ_{IG}	0.3264	(0.5295)			

distributed disturbances of the system equations facilitating the derivation of closed-form expressions for the objective function. A violation of this assumption leads to inefficient parameter estimates and could indicate, that we did not succeed in incorporating all dependencies adequately by choosing the quadratic polynomial specification. So, as an additional diagnostic check for this final specification, consider the sample correlation between the three residual series

$$\hat{\rho} = \begin{bmatrix} 1 & -0.0221 & -0.0096 \\ . & 1 & -0.0046 \\ . & . & 1 \end{bmatrix}.$$

Compared to the sample correlations for the equation–by–equation residuals reported earlier, these are obviously much closer to zero and generally insignificant. The three scatter plots for the pairwise realized cdf's for the system residuals in Figure 4.9 now also appear uniformly distributed over the entire range, indicating that the quadratic polynomials have successfully accounted for the non–linear contemporaneous dependencies observed in the equation-b-y–equation residuals and that our assumptions about the innovation distributions are empirically accurate.[12]

[12] Also, the system counterparts to Figures 4.5 to 4.7, i.e., the scatter plots of the pairwise residuals and the corresponding estimated quadratic polynomials and kernel based estimators, do not reveal any neglected non–linear dependencies. For example, the parameter estimates for the quadratic polynomial involving the Bipower variation residuals as a function of the shocks to the return equation equal 0.1060 (0.0218), -0.0269 (0.0190) and -0.0035 (0.0150), respectively, none of which are significant at conventional levels.

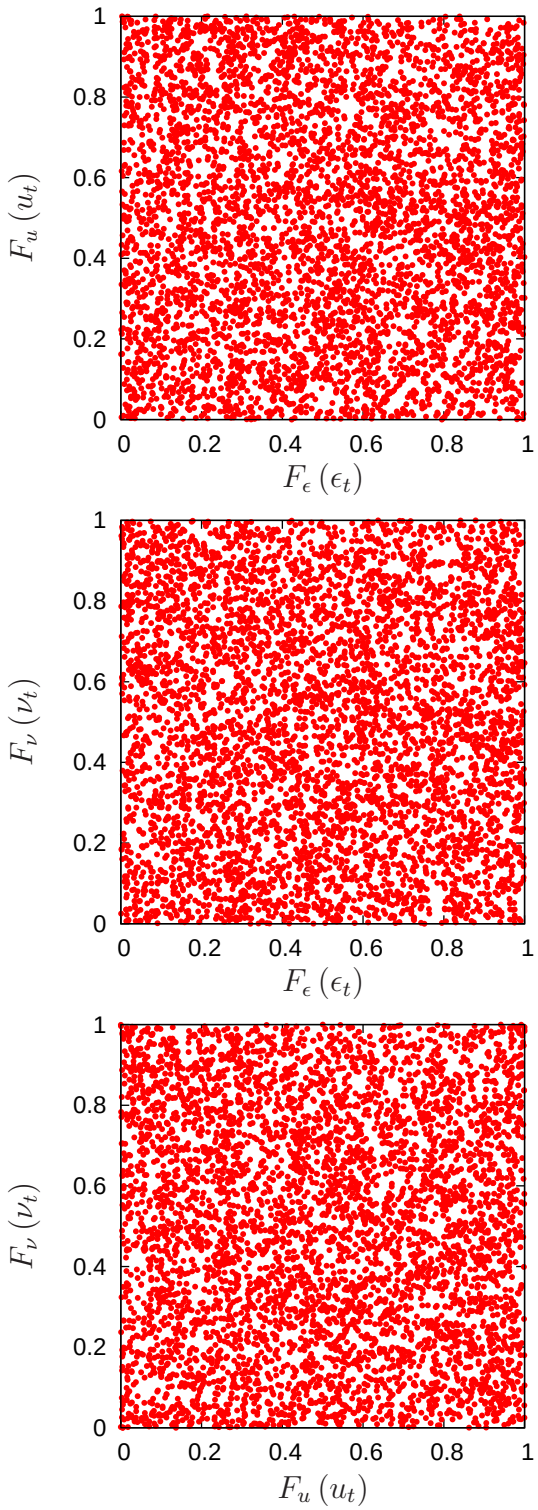

Figure 4.9: Realized CDF scatter plot of the system innovations.

4.4 Model Simulations

The discussion in the previous section suggests that the model performs an exemplary job in terms of describing the one–day–ahead conditional transition densities when judged by the standard maximum likelihood criteria and different model diagnostics. Meanwhile, in order to better understand the workings and possible limitations of a given model, it is often instructive to consider its ability to account for other aspects of the data through the use of simulations. To this end, we generate 105,040 observations from the estimated system, keeping only the last 5,040 observations corresponding to the sample size of our data; i.e., the first 100,000 simulated observations serve as a large burn–in period. We then repeat this 25,000 times, leaving us with 25,000 simulated "daily" sample paths for the returns, logarithmic Bipower variation, and relative jump series. To illustrate, Figure 4.10 shows one such representative set of simulated data. The basic similarities for each of the series with those of the original data in Figure 2.3 are striking, and indeed shows the model to be broadly consistent with the data.

More formally, consider the summary statistics in Table 2.1. By calculating the same set of summary statistics for each of the 25,000 simulated sample paths, we obtain a model–implied, i.e. bootstrapped, sample distribution for the respective statistics. If the model provides an adequate description of the observed data, the realized values of the corresponding sample statistics should lie within reasonable confidence intervals, say 95%, of the bootstrapped sample distributions. Table 4.4 provides these 95% simulated confidence intervals for the standard set of summary statistics, as well as the actual sample values from Table 2.1. We also report actual and simulated quantiles for each of the series, and illustrate these in Figure 4.11. Nearly all of the sample statistics, including all of the reported 0.01 to 0.99 quantiles, lie within the simulated confidence bands. Only the realized skewness and kurtosis for the returns and the realized kurtosis for the logarithmic Bipower variation (with the latter probably spilling over to that of the logarithmic realized variance) fall outside the 95% bands. Although our maximum likelihood based inference does not seem to favor this, this could presumable be "fixed" by allowing for a leptokurtic skewed error distribution in the return equation, either parametrically or through the uses of more flexible semi–non–parametric density estimation as in, e.g., Gallant and Nychka (1987) and Gallant and Tauchen (1989). Importantly, however, the distributional properties of the jumps are well captured by our model.

Exploring the dynamic implications of the model, Figure 4.12 shows the sample autocorrelations and partial autocorrelations with the corresponding simulated 95% confidence bands. As can be seen from the figure, the observed short–run dynamics of both the returns and the relative jump series are generally consistent with those of our model. Meanwhile, the HAR model for $\log BV_t$, as well as the model's implications for $\log RV_t$, both fall somewhat short in terms of reproducing the highly significant and very slowly decaying sample autocorrelations over longer multi–month lags. One reason might be, that the HAR model is formally no long memory model and the number and horizons of the different components have

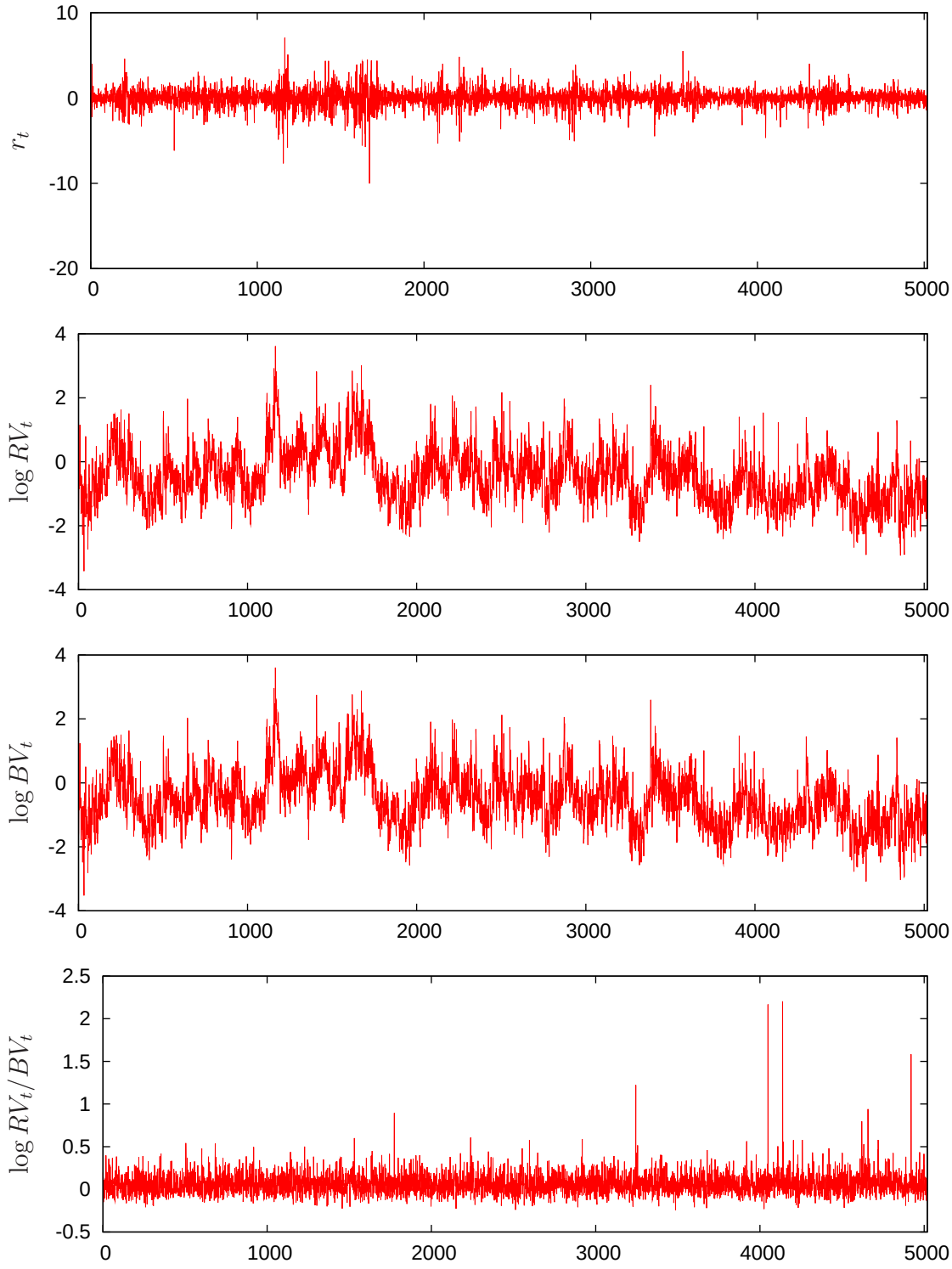

Figure 4.10: Simulated paths of returns, logarithmic realized variance, logarithmic Bipower variation and jumps.

Table 4.4: Simulation Results

	r_t		$\log RV_t$		$\log BV_t$		$\log\left(\frac{RV_t}{BV_t}\right)$	
stat.	realized	95% intervals	realized	95% intervals	realized	95% intervals	realized	95% intervals
Mean	0.0254	(-0.0125,0.0416)	-0.5139	(-0.7501,-0.3148)	-0.5817	(-0.8159,-0.3788)	0.0678	(0.0611,0.0687)
Std.Dev.	1.0946	(0.8539,1.1504)	0.8775	(0.7382,0.9312)	0.8845	(0.7447,0.9377)	0.1263	(0.1200,0.1354)
Skew.	-2.1648	(-1.8996,0.0906)	0.5948	(-0.0211,0.5765)	0.5416	(-0.0252,0.5711)	1.7761	(0.9838,3.8142)
Exc.Kurt.	96.2483	(3.2207,37.8147)	1.7981	(-0.0476,1.2418)	1.4807	(-0.0354,1.2464)	12.2675	(3.1382,60.7865)
$Q_{0.01}$	-2.6479	(-3.4341,-2.3966)	-2.3275	(-2.7117,-2.0456)	-2.3868	(-2.7956,-2.1305)	-0.1517	(-0.1720,-0.1548)
$Q_{0.025}$	-2.0527	(-2.4330,-1.7798)	-2.0632	(-2.3743,-1.7998)	-2.1377	(-2.4537,-1.8780)	-0.1303	(-0.1394,-0.1268)
$Q_{0.05}$	-1.5535	(-1.7945,-1.3384)	-1.8321	(-2.0994,-1.5834)	-1.9172	(-2.1746,-1.6577)	-0.1027	(-0.1116,-0.1013)
$Q_{0.10}$	-1.0895	(-1.2252,-0.9229)	-1.5848	(-1.7958,-1.3244)	-1.6667	(-1.8696,-1.3969)	-0.0737	(-0.0792,-0.0703)
$Q_{0.25}$	-0.4626	(-0.5275,-0.3872)	-1.1147	(-1.3135,-0.8754)	-1.1859	(-1.3812,-0.9436)	-0.0143	(-0.0225,-0.0147)
$Q_{0.50}$	0.0511	(0.0308,0.0767)	-0.5527	(-0.7797,-0.3452)	-0.6163	(-0.8443,-0.4098)	0.0538	(0.0470,0.0551)
$Q_{0.75}$	0.5446	(0.4790,0.5901)	0.0173	(-0.2421,0.2279)	-0.0533	(-0.3044,0.1683)	0.1339	(0.1264,0.1366)
$Q_{0.90}$	1.0964	(0.9484,1.1850)	0.6022	(0.2553,0.8023)	0.5372	(0.1968,0.7448)	0.2218	(0.2091,0.2242)
$Q_{0.95}$	1.5001	(1.2961,1.6478)	0.9853	(0.5634,1.1937)	0.9352	(0.5084,1.1375)	0.2799	(0.2656,0.2870)
$Q_{0.975}$	1.9627	(1.6544,2.1554)	1.3462	(0.8420,1.5616)	1.2815	(0.7860,1.5113)	0.3371	(0.3211,0.3524)
$Q_{0.99}$	2.5991	(2.1481,2.9251)	1.8250	(1.1784,2.0496)	1.8240	(1.1285,2.0040)	0.4322	(0.3964,0.4557)

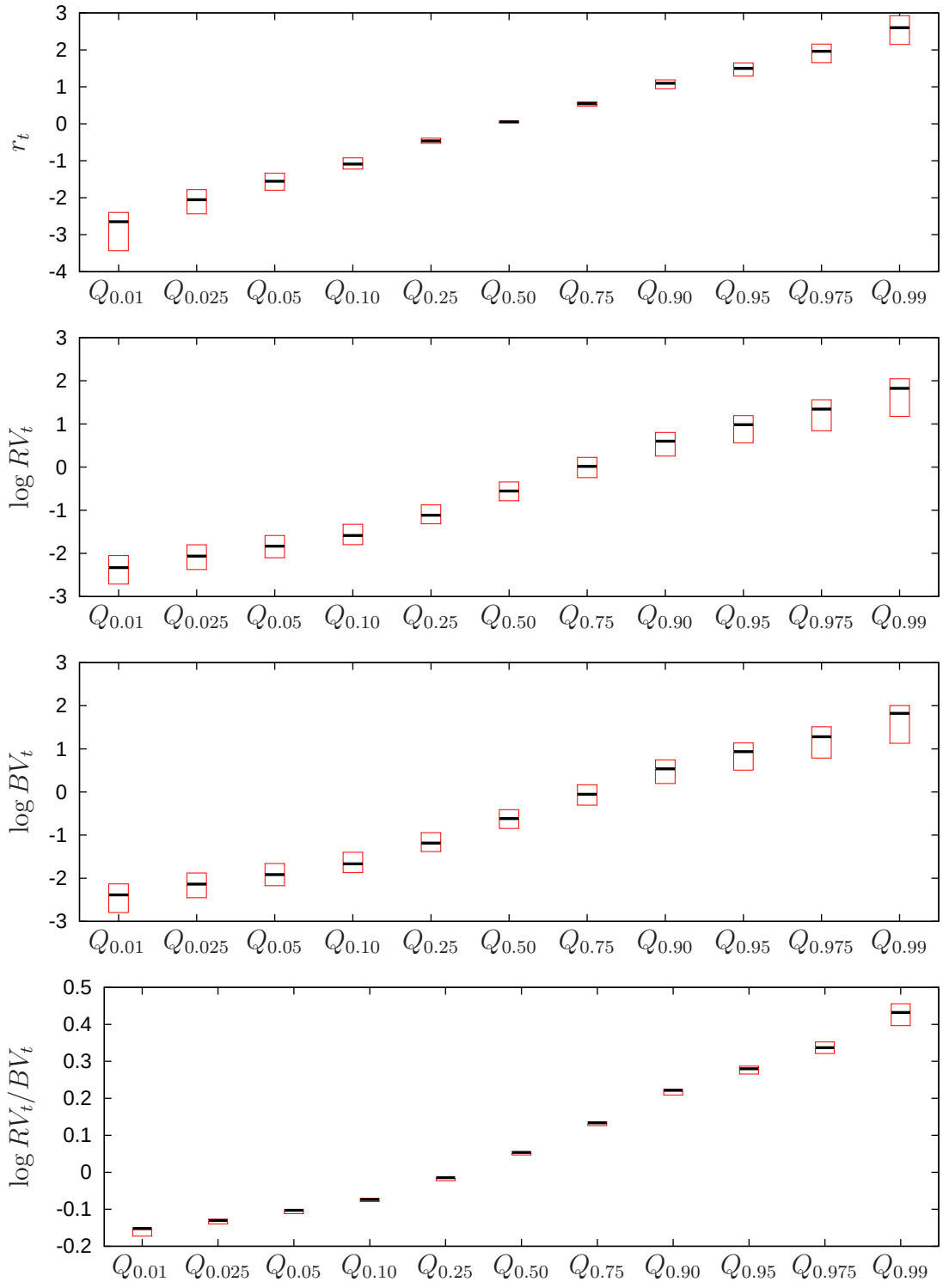

Figure 4.11: Sample quantiles of returns, logarithmic realized variance, logarithmic Bipower variation and jumps with 95% simulated confidence intervals.

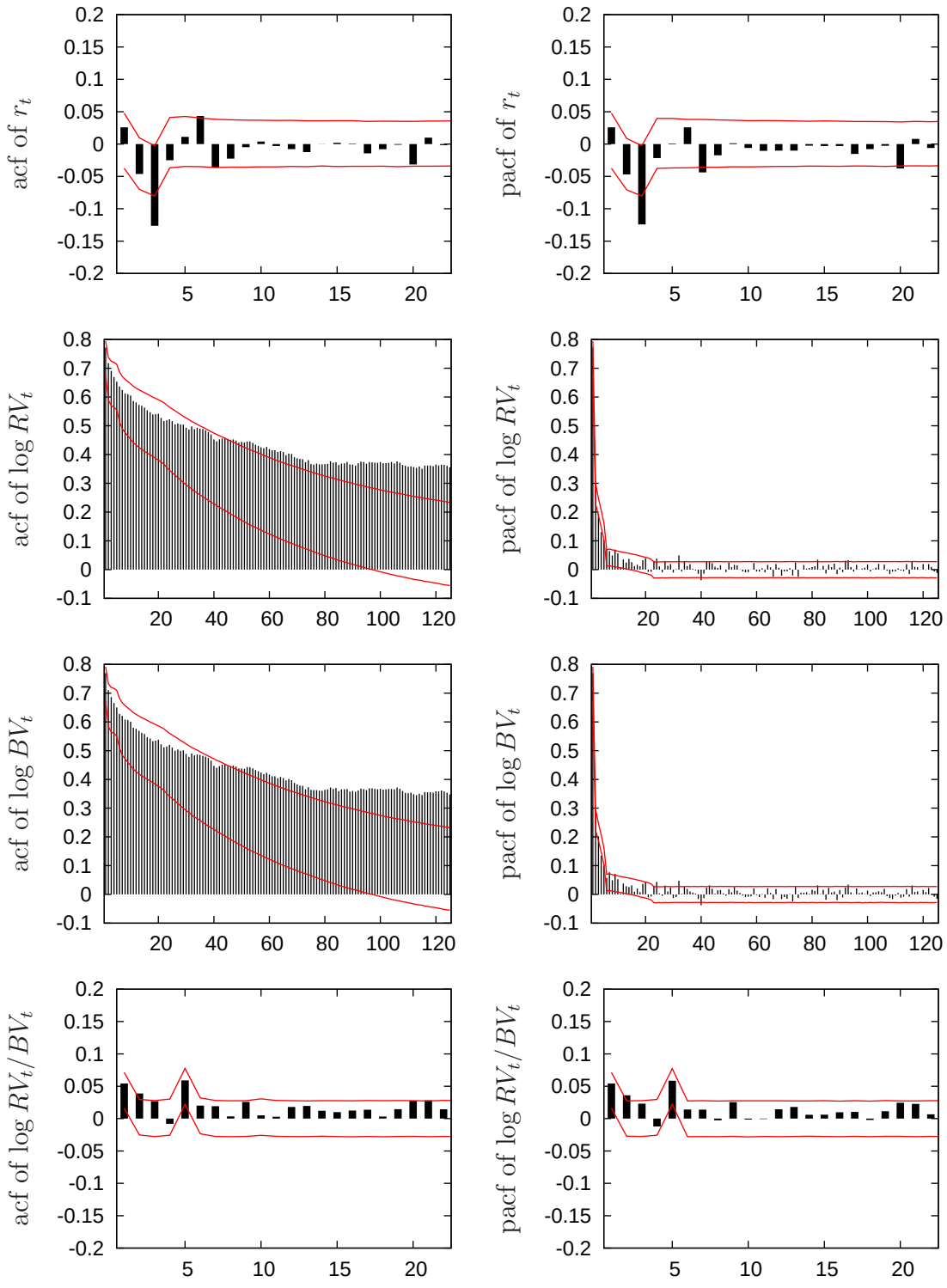

Figure 4.12: Sample autocorrelations and partial autocorrelations of returns, logarithmic realized variance, logarithmic Bipower variation and jumps. The dashed lines give the upper and lower ranges of the simulated 95% confidence intervals.

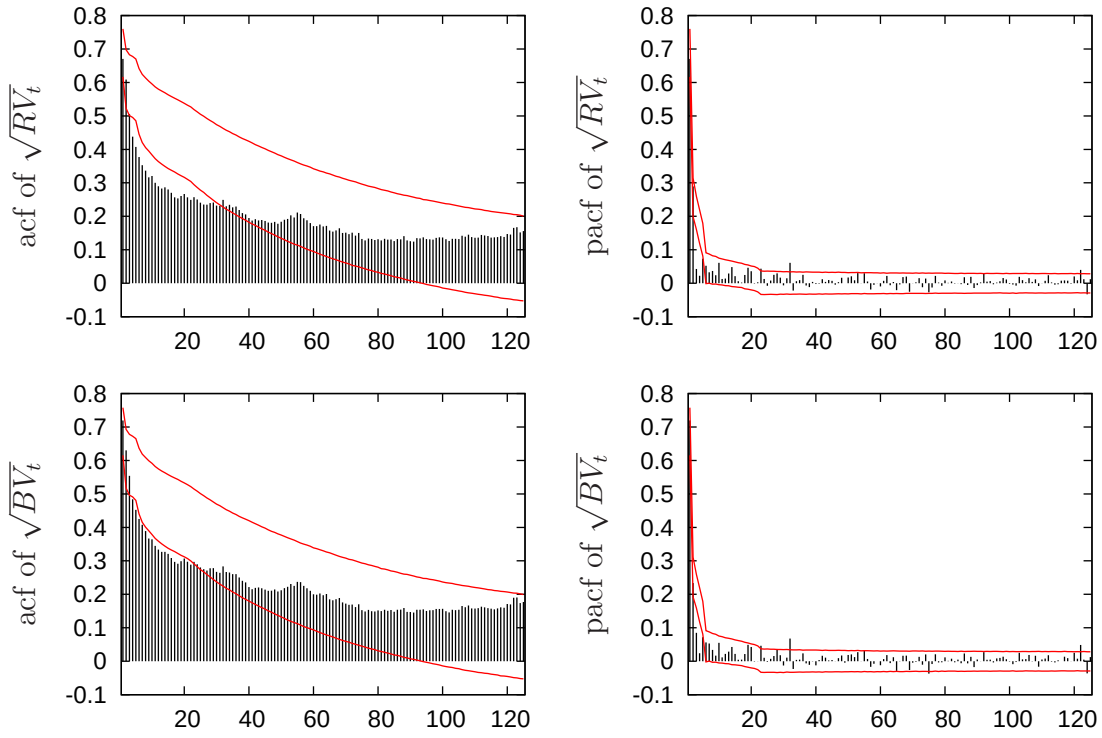

Figure 4.13: Sample autocorrelations and partial autocorrelations of realized volatility and Bipower variation in standard deviation form. The dashed lines give the upper and lower ranges of the simulated 95% confidence intervals.

to be adjusted correspondingly. As previously noted, the inclusion of quarterly or longer–run realized variation measures on the right–hand–side of the HAR model for $\log BV_t$ would presumable remedy this deficiency (see also the simulations reported in Corsi (2004), which shows that HAR models with longer lags can get remarkably close to reproducing the autocorrelations of true long–memory volatility processes.) At the same time, however, Figure 4.13 shows that the autocorrelations for the Bipower and realized variation expressed in standard deviation form, as would be of interest in many practical applications, both are well accounted for by our relatively simple and easy–to–implement final preferred model.

4.5 Conclusion

Motivated by the recent empirical results on the relevance of jumps to total price variation derived from high–frequency based realized volatility and Bipower variation measures, this chapter developed a joint discrete–time model for returns and volatility by explicitly disentangling the dynamics of the continuous volatility and jump components. We show that the often observed leverage effect, or asymmetry in the lagged return volatility relationship, primarily acts through the continuous volatility component. Moreover, our modeling approach also facilitates the closer examination of contemporaneous inter–dependencies providing results that do not only reveal the statistical importance of such interrelations, but also point towards a similar mechanism in the contemporaneous leverage effect. Our findings are thus in line with most of the parametric continuous–time jump diffusion models employed in the literature, which typically introduce the leverage effect by correlating the Brownian motions driving the return and continuous volatility processes.

Also, the simulation study demonstrates the adequacy of our model specification. Importantly, we are able to capture the dynamic patterns of the jumps and square root Bipower variation and can reproduce most of the distributional properties using a standard normal distribution, a normal–mixture and NIG–IG mixture for returns, the continuous volatility component and the jumps, respectively. Moreover, our specification seems to be quite robust as we have repeated all estimations and computations presented here for a shorter sample ranging from 1990 to 2002 — a sample period that excludes the largest price movements such as the 1987 crash — yielding similar results.

Our modeling strategy followed here has several advantages over some of the other reduced–form realized volatility and GARCH modeling procedures recently considered in the literature (see also Chapter 3).

Firstly, by explicitly disentangling the dynamics of the two volatility components, our results clearly show that the jumps are much less persistent, and hence less predictable, than the continuous sample path variation. This in turn should result in improved volatility forecasts, with direct and important implications for interval forecasts and corresponding risk management decisions. Indeed, it would be interesting to further explore this conjecture by directly comparing the model

forecasts and risk measurements from the model developed here with those obtained from other popular volatility forecasting procedures, including GARCH type models and simpler reduced form realized volatility models that do not explicitly differentiate between the continuous and jump components. Along these lines, if the two volatility components carry different risk premia, separately modeling and forecasting each of the components, should also result in more accurate prices for options and other derivative instruments.

Secondly, by providing a highly accurate description of the discrete–time joint dynamics of the returns and the two volatility components, our model can be used in the indirect estimation of other parametric volatility models, effectively incorporating the information contained in the high–frequency data.[13] More specifically, the flexibility and recursive structure of the model coupled with the ready availability of its analytic derivatives, combine to make it an ideal candidate for the role of an auxiliary model, or score generator, within the Efficient Method of Moment (EMM) estimation framework of Gallant and Tauchen (1996), or the General Scientifc Modeling (GSM) approach developed by Gallant and McCulloch (2005). In particular, even though the discrete–time model has no direct continuous–time analog, and may in fact be consistent with many different continuous time formulations, it is nonetheless highly informative about the general features that need to be accounted for in the data. Relying on the likelihood function from the discrete–time model as a summary of the data, thus facilitates the estimation and empirical assessment of much richer Poisson jump diffusion or more general Lévy–driven continuous–time stochastic volatility models than hitherto considered in the literature by affording a numerical evaluation of the model likelihoods through the use of long artificial simulations. In Bollerslev et al. (2006a) we have found some encouraging results along these lines based on the discrete time model developed here and the GSM approach.

[13] High–frequency data based non–parametric realized volatility measures have previously been used in the estimation of parametric continuous time stochastic volatility models by Barndorff-Nielsen and Shephard (2002a) and Bollerslev and Zhou (2002).

5 A Multivariate Generalized Hyperbolic Stochastic Volatility Model

So far we have primarily focused on exploiting the high–frequency information inherent in the realized variation measures for modeling the dynamics of individual assets, or more specifically of the S&P500 index futures. However, for financial institutions the knowledge of the individual as well as the cross–sectional dynamics of multiple assets is of major importance. We therefore turn now to the multivariate modeling of asset returns using the realized covariance measure. Note that at least to our knowledge the number of existing multivariate realized covariance models is quite limited. In fact we are aware of only five papers addressing this issue. While Bauer and Vorkink (2007) and Da and Schaumburg (2006) mainly consider factor models for the realized covariance, Andersen et al. (2001b), Chiriac and Voev (2007) employ VARFIMA–type models, and Voev (2007) considers various ARMAX models as well as a shrinkage realized covariance forecast. However, they all assume that the "true" covariance matrix is observable. Moreover, they solely focus on the dynamics of the realized covariance matrix. Making also use of the realized covariance measure, we instead model jointly the dynamics of multiple asset returns and the unknown covariance matrix.

In particular, in this chapter we provide a multivariate generalization of the normal inverse Gaussian stochastic volatility model originally introduced by Barndorff-Nielsen (1997). This modeling framework is especially appealing as it combines stochastic volatility features with those of GARCH models by specifying the mean of the stochastic volatility to depend on past observations, such as lagged squared returns. Moreover, the model is designed to capture most of the stylized facts of stock returns. E. g. the fat–tailedness of the return distribution is modeled by normal inverse Gaussian or by the more general class of generalized hyperbolic distributions, which nests the normal invers Gaussian as a special case. Note that the usefulness of this class of distributions for univariate financial modeling has also been considered in e.g. Barndorff-Nielsen (1995), Eberlein and Keller (1995), Eberlein et al. (1998) and Küchler et al. (1999). However, while these approaches assume i.i.d. innovations the model of Barndorff-Nielsen (1997) explicitly accounts for the observed volatility clustering as is discussed below.

A more general representation of this type of model is given by the following

univariate generalized hyperbolic stochastic volatility model

$$r_t = \mu + \beta\sigma_t^2 + \sigma_t\epsilon_t \quad \text{with} \quad \epsilon_t \overset{i.i.d.}{\sim} N(0,1) \tag{5.1}$$

$$\sigma_t^2 \,|\mathcal{X}_{t-1} \overset{i.i.d.}{\sim} GIG(\lambda, \delta, \sqrt{\alpha^2 - \beta^2}) \tag{5.2}$$

$$E(\sigma_t^2 \,|\mathcal{X}_{t-1}) = h_t(\mathcal{X}_{t-1}; \theta), \tag{5.3}$$

where $\mathcal{X}_{t-1}$ subsumes past return observations and $h_t(\mathcal{X}_{t-1}; \theta)$ is a deterministic parametric function. The model specification is primarily motivated by the mixture–of–distributions hypothesis (see Clark (1973)), that states that returns are conditionally normal distributed, but with a latent stochastic variance. In particular, the return distribution is given by a normal variance–mean mixture implying a return equation that is quite standard in the GARCH–in–mean and stochastic volatility–in–mean literature, and, thus, captures important stylized facts.[1] More precisely, the mean mixing introduces the volatility into the return equation and, thus, accounts for the risk premium resulting also in a skewed distribution. The variance mixing, in turn, accounts for heavier tails relative to the normal distribution. Here, the variance and mean mixing variable is assumed to be generalized inverse Gaussian distributed resulting in generalized hyperbolic distributed returns. The generalized hyperbolic distribution was introduced by Barndorff-Nielsen (1977) and turned out to be very flexible allowing also for heavier than log linear tails as is usually observed for asset returns. In fact, most of the distributions commonly used in the finance literature are nested in the generalized hyperbolic distribution or can be obtained as limiting distributions, such as for example the t– or NIG distributions (see e.g. Eberlein and von Hammerstein (2004)).

While Taylor (1986) introduced the volatility clustering effect into the mixture–of–distributions setup by specifying a Gaussian autoregression for the stochastic latent variance, the generalized hyperbolic stochastic volatility model incorporates this feature by specifying the mean of the stochastic latent variance as a function of past return observations, see equation (5.3).[2] As such it combines stochastic volatility with GARCH features, which is the main characteristic of this model. Assuming a normal inverse Gaussian distribution, i.e. $\lambda = -1/2$, with $\mu = \beta = 0$, Barndorff-Nielsen (1997), for example, suggested an ARCH–type specification, i.e. $h_t(\mathcal{X}_{t-1}; \theta) = \omega_0 + \sum_{i=1}^{p} \omega_i r_{t-1}^2$, whereas Andersson (2001) and Jensen and Lunde (2001) considered GARCH and asymmetric power ARCH specifications, respectively.

From the viewpoint of modeling the dynamics of stock prices using volatility measures, the generalized hyperbolic stochastic volatility model is very appealing

[1] The GARCH in mean model was introduced by Engle et al. (1987) and applied in for example Campbell and Hentschel (1992) and Shin (2005). Koopman and Uspensky (2002) were the first to consider a risk premium effect within a discrete–time stochastic volatility model.

[2] Note that we model $\mathrm{E}\left[\sigma_t^2 | \mathcal{X}_{t-1}\right]$ as a function of past observables rather than the parameter δ of the GIG distributed variable, i.e. $\delta = h_t(\mathcal{X}_{t-1}; \theta)$, as originally suggested in Barndorff-Nielsen (1997). In the multivariate extension our specification turns out to be less restrictive as it allows to model dynamic conditional correlations. However, note that in the univariate cases considered in Barndorff-Nielsen (1997), i.e. for $\mu = \beta = 0$, both representations are equivalent.

as it possesses a nice interpretation. Intuitively, the model acknowledges the fact, that volatility measures, such as daily squared returns or realized volatility, are unbiased, but importantly noisy measures of the latent volatility. As such the approach differs from most of the existing realized volatility models, which build on the assumption that the unknown volatility is perfectly observable via the realized volatility (see e.g. our trivariate model of Chapter 4, as well as Andersen et al. (2003), Forsberg and Bollerslev (2002), and Giot and Laurent (2004), where returns are modeled by equating the conditional variance with realized variance). Instead, the generalized hyperbolic stochastic volatility model still treats the volatility as being unknown while recognizing that the (realized) volatility measures provide important information on the true volatility process. This information is exploited by allowing the mean of the stochastic volatility to depend on past observations of the volatility measures.[3] Moreover, the backward looking nature of this particular specification does not only account for the fact, that the volatility measures are available only ex post, but also facilitates the computation of return and volatility forecasts.

So far, the normal inverse Gaussian stochastic volatility model was primarily used for modeling the return dynamics of a single asset. However, for portfolio optimization and risk management decisions the knowledge of the joint distribution and dynamics of multiple assets is crucial. With the recent availability of the realized covariance measure the generalized hyperbolic model becomes a natural candidate for modeling the joint dynamics of asset returns. In particular, applying the ideas of the generalized hyperbolic stochastic volatility model to the multivariate case implies that the full covariance matrix is modeled directly. This is in contrast to most of the standard multivariate stochastic volatility models (see e.g. Chib et al. (2006), Harvey et al. (1994) and Jacquier et al. (1994), in which the time–variation of the covariance matrix is introduced by stochastic specifications for the vector of (logarithmic) volatilities and a constant correlation matrix of the return innovations. Alternative approaches that model also the full covariance matrix directly were proposed in for example Asai et al. (2006), Gourieroux (2006), and Philipov and Glickman (2006). However, these studies disregard the potentially valuable information contained in the high–frequency based measures by assuming a purely latent stochastic covariance process. Instead, in the multivariate generalized hyperbolic model this information is exploited by specifying the mean of the latent stochastic volatility as a function of past observations of the realized covariance matrix. Importantly, note also that based on its definition the realized covariance matrix is positive definite (as long as the number of assets does not exceed the sampling frequency) and, thus, can straightforwardly be used for modeling the unknown covariance matrix. As such, the model additionally allows to consider generally higher dimensional portfolios as would be the case with daily data.[4]

[3] As such the model tackles the errors–in–variable problem of the (realized) volatility measures. This problem has also been noted by e.g. Andersen et al. (2005).

[4] Obviously, if only daily data would be available, then the empirical covariance measure to be

The remainder of this chapter is organized as follows. The next section discusses briefly the relationship between the univariate generalized hyperbolic stochastic volatility and GARCH–type models. Section 5.2 presents important and useful properties of the multivariate generalized hyperbolic distribution. Section 5.3 then introduces the general setup of our multivariate model along with two particular specifications for the mean function that imply time–varying volatilities and correlations. In Section 5.4 we focus on the normal inverse Gaussian case and apply our model to the three component stocks of the S&P500 index discussed in Section 2.2.2. Section 5.5 concludes.

5.1 The Relation between the Generalized Hyperbolic Stochastic Volatility Model and GARCH Models

In the previous discussion we have seen that the generalized hyperbolic stochastic volatility model combines stochastic volatility features with those of GARCH models. One may therefore be interested in how the model is related to the existing stochastic volatility literature and to GARCH models. It is well–known that, generally, the main difference between these two popular modeling approaches is given by their volatility specification. In particular, while the GARCH models assume that financial volatility is a deterministic function of past returns, the stochastic volatility models specify an extra stochastic process for the volatility. The standard stochastic volatility models moreover assume that the volatility is latent and independent of returns (or other observables), implying also that the innovations in the return and volatility equations are assumed to be independent. However, these assumptions are oftentimes relaxed, especially if the leverage effect is incorporated.[5]

Given that the generalized hyperbolic stochastic volatility model specifies a latent stochastic volatility whose mean only is allowed to dependent on past observables, its familiarity with the stochastic volatility models is obvious. However, based on the GARCH–type specification of the mean function, it is also of interest to assess its relation to the traditional GARCH models.

used in the model could be constructed over longer horizons such as a month. Note that such a switch from the daily to the weekly or monthly level is common practice in portfolio optimization applications and can also be used here to allow for even larger portfolio dimensions exceeding the intraday sampling frequency. In this context, it is important to emphasize here once more, that the realized covariance matrix is not guaranteed to be always positive definite "by construction" as is sometimes assumed in the literature (see e.g. Bauer and Vorkink (2007)).

[5]See for example the models of Asai and McAleer (2005, 2006) and Danielsson (1994), in which the volatility depends on lagged absolute returns, or Chan et al. (2006), Harvey and Spephard (1996), and Yu (2005), who allow for correlations among the volatility and return innovations. Similar leverage specifications have also been introduced in the continuous–time stochastic volatility models by correlating the Brownian motions of the return and volatility processes, see e.g. Bates (2000), Eraker et al. (2003) and Pan (2002).

To this end, we first review some properties of the generalized hyperbolic distribution. As noted previously, the generalized hyperbolic distribution can be derived as a normal variance–mean mixture. In particular, let $X = \mu + \beta Z + \sqrt{Z}Y$ with $Y \overset{i.i.d.}{\sim} N(0,1)$ and Z denoting a positive random mixing variable, that is $GIG(\lambda, \delta, \sqrt{\alpha^2 - \beta^2})$ distributed with probability density function

$$f_{GIG}(z; \lambda, \delta, \gamma) = \frac{(\gamma/\delta)^\lambda}{2K_\lambda(\delta\gamma)} z^{\lambda-1} \exp\left(-\frac{1}{2}\left(\delta^2 z^{-1} + \gamma^2 z\right)\right), \tag{5.4}$$

where $\gamma = \sqrt{\alpha^2 - \beta^2}$ and $K_\lambda(\cdot)$ is the modified Bessel function of the third kind and of order $\lambda \in \mathbb{R}$. The parameters of the GIG distribution are restricted by

$$\begin{array}{ll} \delta \geq 0 \text{ and } |\beta| < \alpha & \text{if } \lambda > 0 \\ \delta > 0 \text{ and } |\beta| < \alpha & \text{if } \lambda = 0 \\ \delta > 0 \text{ and } |\beta| \leq \alpha & \text{if } \lambda < 0, \end{array}$$

and the mean and variance of Z are given as

$$\begin{aligned} \mathrm{E}[Z] &= \left(\frac{\delta}{\gamma}\right) \frac{K_{\lambda+1}(\delta\gamma)}{K_\lambda(\delta\gamma)} \\ \mathrm{V}[Z] &= \left(\frac{\delta}{\gamma}\right)^2 \left(\frac{K_{\lambda+2}(\delta\gamma) - K^2_{\lambda+1}(\delta\gamma)}{K_\lambda(\delta\gamma)}\right). \end{aligned}$$

Based on this particular normal mean–variance–mixture, the resulting distribution of X is a generalized hyperbolic distribution with probability density function given by

$$\begin{aligned} f_{GH}(x; \lambda, \alpha, \beta, \delta, \mu) &= \frac{(\delta\gamma)^\lambda (\delta\alpha)^{1/2-\lambda}}{\sqrt{2\pi}\delta K_\lambda(\delta\gamma)} \left(1 + \frac{(x-\mu)^2}{\delta^2}\right)^{\lambda/2-1/4} \\ &\quad \times K_{\lambda-1/2}\left(\alpha\delta\sqrt{1 + \frac{(x-\mu)^2}{\delta^2}}\right) \exp\left(\beta(x-\mu)\right) \end{aligned}$$

with $\mu \in \mathbb{R}$ and the parameter restrictions induced by the GIG distribution (see also Barndorff-Nielsen (1977) and Barndorff-Nielsen and Blæsild (1981)). Mean and variance of X are available in closed form and are given by

$$\begin{aligned} \mathrm{E}[X] &= \mu + \beta \mathrm{E}[Z] \qquad (5.5) \\ &= \mu + \frac{\beta\delta}{\gamma} \frac{K_{\lambda+1}(\delta\gamma)}{K_\lambda(\delta\gamma)} \\ \mathrm{V}[X] &= \mathrm{E}[Z] + \beta^2 \mathrm{V}[Z] \qquad (5.6) \\ &= \frac{\delta}{\gamma} \frac{K_{\lambda+1}(\delta\gamma)}{K_\lambda(\delta\gamma)} \\ &\quad + \frac{\beta^2\delta^2}{\gamma^2} \left(\frac{K_{\lambda+2}(\delta\gamma)}{K_\lambda(\delta\gamma)} - \left(\frac{K_{\lambda+1}(\delta\gamma)}{K_\lambda(\delta\gamma)}\right)^2\right). \end{aligned}$$

So, obviously mean and variance depend on the first, or first two moments of the GIG distributed random variable, respectively.

Now, following Barndorff-Nielsen (1977) and Andersson (2001) we assume that $\mu = \beta = 0$. In this case the mean and variance of a symmetric generalized hyperbolical distributed random variable $X \sim GH(\lambda, \alpha, 0, \delta, 0)$ reduce to

$$\begin{aligned} \mathrm{E}[X] &= 0 \\ \mathrm{V}[X] &= \mathrm{E}[Z] \\ &= \left(\frac{\delta}{\alpha}\right) \frac{K_{\lambda+1}(\delta\alpha)}{K_\lambda(\delta\alpha)}. \end{aligned}$$

In the generalized hyperbolic stochastic volatility model this implies that the returns are conditionally generalized hyperbolic distributed, with

$$\begin{aligned} \mathrm{E}[r_t | \mathcal{X}_{t-1}] &= 0 \\ \mathrm{V}[r_t | \mathcal{X}_{t-1}] &= h_t(\mathcal{X}_{t-1}; \theta). \end{aligned}$$

Obviously, this restriction only holds if at least one parameter of the generalized inverse Gaussian distribution is time–varying. Oftentimes this is the parameter δ, see e.g. Barndorff-Nielsen (1977). Moreover, note that if $h_t(\mathcal{X}_{t-1}; \theta) = \omega_0 + \sum_{i=1}^{p} \omega_i r_{t-i}^2 + \sum_{j=1}^{q} \psi_j \sigma_{t-j}^2$, then the generalized hyperbolic stochastic volatility model implies a conditional variance that is simply given by a GARCH(p, q) specification.

Alternatively, let us consider the GARCH(p, q) model generally defined by

$$r_t = \sqrt{h_t(\mathcal{X}_{t-1}; \theta)}\eta_t \tag{5.7}$$

$$h_t(\mathcal{X}_{t-1}; \theta) = \omega_0 + \sum_{i=1}^{p} \omega_i r_{t-i}^2 + \sum_{j=1}^{q} \psi_j h_{t-j}, \tag{5.8}$$

with i.i.d. innovations η_t having zero mean and unit variance, such that $\mathrm{E}[r_t|\mathcal{X}_{t-1}] = 0$ and $\mathrm{V}[r_t|\mathcal{X}_{t-1}] = h_t(\mathcal{X}_{t-1}; \theta)$. In the following we assume that the return innovations are generalized hyperbolic distributed. Then, by the scaling property of the generalized hyperbolic distribution, the conditional returns are distributed according to a generalized hyperbolic distribution with mean zero and variance given by

$$\mathrm{V}[r_t | \mathcal{X}_{t-1}] = \omega_0 + \sum_{i=1}^{p} \omega_i r_{t-i}^2 + \sum_{j=1}^{q} \psi_j \sigma_{t-j}^2.$$

Hence, the generalized hyperbolic GARCH model imposes the same restrictions on the parameter space of the generalized hyperbolic distribution as the generalized hyperbolic stochastic volatility model. So, for $\mu = \beta = 0$ the generalized hyperbolic stochastic volatility model as well as the generalized hyperbolic GARCH model are equivalent as they both imply the same distribution for the conditional returns.

Intuitively, this means that the GARCH model incorporates the uncertainty of the conditional variance by assuming a normal variance mixture distribution for the return innovations, which generally exhibits heavier tails than the Gaussian distribution. Note that the equivalence also holds for nonzero location parameter μ yielding in both models the above (similarly restricted) generalized hyperbolic distribution but with $\mathrm{E}[r_t|\mathcal{X}_{t-1}] = \mu$.[6]

This is an interesting result, as it bridges the gap between stochastic volatility and GARCH models. In fact, many of the existing GARCH models are nested in the generalized hyperbolic GARCH model and as such can also be interpreted as Gaussian–based, i.e. based on a normal–variance mixture, stochastic volatility models. A popular example is the GARCH model with t–distributed innovations proposed by Bollerslev (1987). This model is equivalent to a Gaussian–based stochastic volatility model with conditional variance being distributed according to a reciprocal Gamma distribution with mean given by a GARCH specification.

Noteworthy, however, the equivalence is no longer valid, if an asymmetric generalized hyperbolic distribution is considered, i.e. β is nonzero. In this case, the term $\beta^2\mathrm{V}[Z]$ in equation (5.6) does not vanish and, thus, imposes other restrictions than those implied by the GARCH model, which are the same as before.[7] One may conjecture that the two models coincide again if $\mathrm{V}[Z] = 0$, however, in this case, the implied conditional return distribution of the stochastic volatility model is Gaussian. This illustrates nicely that the stochastic volatility model only differs from the Gaussian GARCH model by assuming a nondegenerate distribution of the conditional variance. This has also been noted in Barndorff-Nielsen (1997). As such, the generalized hyperbolic stochastic volatility model can be used to test for the validity of the Gaussian GARCH model by testing whether $\mathrm{V}[Z] = 0$.

Note that the derived results can be generalized to any stochastic volatility model of the normal variance–mean mixture type as given in (5.1) to (5.3), if it has a positive mixing distribution, that possesses the scaling property, and if the corresponding GARCH model has return innovations that are distributed according to the corresponding mixture distribution.

5.2 Properties of the Multivariate Generalized Hyperbolic Distribution

Originally, the generalized hyperbolic distribution was introduced by Barndorff-Nielsen (1977) for modeling the distribution of the size of sand particles. However, exhibiting semiheavy tails and a closed–form expression for the probability density function this distribution has also gained much attention in the finance literature,

[6] This is easily verified if the return equation in the GARCH model is written as $r_t = \mu + \sigma_t\eta_t$ with η_t being distributed as before, and using the fact that the generalized hyperbolic distribution is closed under location and scale transformation.

[7] Obviously, the variance of $r_t|\mathcal{X}_{t-1}$ additionally includes the variance associated with the risk premium, i.e. coming from $\beta\sigma_t^2$ in the return equation.

see e.g. Eberlein and Keller (1995), Eberlein et al. (1998) and Küchler et al. (1999). Moreover, its multivariate version possesses very attractive properties that are especially appealing for modeling the joint distribution of stock returns.

Similarly to the univariate case, the multivariate generalized hyperbolic distribution can be constructed as a multivariate normal variance–mean mixture (see e.g. Barndorff-Nielsen (1977) and Blæsild and Jensen (1981)). In particular, let $\boldsymbol{X}$ now denote a d–dimensional column vector of random variables that is conditionally on the scalar random variable Z multivariate normal distributed with mean vector $\boldsymbol{\mu}+Z\boldsymbol{\beta}\boldsymbol{\Delta}$ and covariance matrix $Z\boldsymbol{\Delta}$, where $\boldsymbol{\mu},\boldsymbol{\beta}\in\mathbb{R}^d$ are $d\times 1$ parameter vectors, and $\Delta\in\mathbb{R}^{d\times d}$ is a positive definite matrix with determinant $|\Delta|=1$ ensuring the identification of the parameters. Moreover, Z is again distributed according to the (univariate) generalized inverse Gaussian distribution, i.e. $GIG(\lambda,\delta,\sqrt{\alpha^2-\boldsymbol{\beta}'\Delta\boldsymbol{\beta}})$ with $\lambda\in\mathbb{R},\delta>0$ and $\alpha^2>\boldsymbol{\beta}'\boldsymbol{\Delta}\boldsymbol{\beta}$. Then, $\boldsymbol{X}$ is multivariate generalized hyperbolic distributed with probability density function given by

$$\begin{aligned} f_{GH_d}(\boldsymbol{x};\lambda,\alpha,\boldsymbol{\beta},\delta,\boldsymbol{\mu},\boldsymbol{\Delta}) &= \frac{(\gamma/\delta)^{\lambda}}{(2\pi)^{d/2}K_{\lambda}(\delta\gamma)} \\ &\quad\times\frac{K_{\lambda-d/2}\left(\alpha\sqrt{\delta^2+(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Delta}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}\right)}{\left(\alpha^{-1}\sqrt{\delta^2+(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Delta}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}\right)^{d/2-\lambda}}\exp\left(\boldsymbol{\beta}'(\boldsymbol{x}-\boldsymbol{\mu})\right) \end{aligned}$$

with $\gamma=\sqrt{\alpha^2-\boldsymbol{\beta}'\Delta\boldsymbol{\beta}}$. Note that popular special cases of this distribution can be obtained by particular choices of λ, i.e. for $\lambda=(d+1)/2$ the multivariate hyperbolic distribution is obtained and for $\lambda=-1/2$ we get the multivariate normal inverse Gaussian distribution.

Obviously, the multivariate generalized hyperbolic distribution belongs to the exponential family. Mean and covariance can therefore easily be derived from the cumulant generating function (see e.g. Blæsild and Jensen (1981)) and are given by

$$\begin{aligned} \mathrm{E}\left[\boldsymbol{X}\right] &= \boldsymbol{\mu}+\frac{\delta}{\gamma}\frac{K_{\lambda+1}(\delta\gamma)}{K_{\lambda}(\delta\gamma)}\boldsymbol{\beta}'\boldsymbol{\Delta} \\ \mathrm{Cov}\left[\boldsymbol{X}\right] &= \frac{\delta}{\gamma}\frac{K_{\lambda+1}(\delta\gamma)}{K_{\lambda}(\delta\gamma)}\boldsymbol{\Delta} \\ &\quad+\frac{\delta^2}{\gamma^2}\frac{\left(K_{\lambda+2}(\delta\gamma)K_{\lambda}(\delta\gamma)-K_{\lambda+1}^2(\delta\gamma)\right)}{K_{\lambda}^2(\delta\gamma)}\boldsymbol{\Delta}'\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\Delta}. \end{aligned}$$

Moreover, suppose that $\boldsymbol{X}$ is a d–dimensional vector of random variables distributed according to the generalized hyperbolic distribution $GH_d(\lambda,\alpha,\boldsymbol{\beta},\delta,\boldsymbol{\mu},\boldsymbol{\Delta})$ and consider a partitioning $(\boldsymbol{X}_1',\boldsymbol{X}_2')'$ of $\boldsymbol{X}$. Let l and k denote the dimensions of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, respectively, and let $(\boldsymbol{\beta}_1',\boldsymbol{\beta}_2')'$ and $(\boldsymbol{\mu}_1',\boldsymbol{\mu}_2')'$ be the corresponding partitions of $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$. The partition of $\boldsymbol{\Delta}$ is

$$\boldsymbol{\Delta}=\begin{pmatrix}\boldsymbol{\Delta}_{11} & \boldsymbol{\Delta}_{12}\\ \boldsymbol{\Delta}_{21} & \boldsymbol{\Delta}_{22}\end{pmatrix}$$

such that $\boldsymbol{\Delta}_{11}$ is a $l \times l$ matrix. Then, Blæsild and Jensen (1981) show that the following three useful properties of the multivariate generalized hyperbolic distribution hold:

(a) The distribution of $\boldsymbol{X}_1$ is the l–dimensional generalized hyperbolic distribution $GH_l(\lambda^*, \alpha^*, \boldsymbol{\beta}^*, \delta^*, \boldsymbol{\mu}^*, \boldsymbol{\Delta}^*)$ with $\lambda^* = \lambda$, $\alpha^* = |\boldsymbol{\Delta}_{11}|^{-1/(2r)} \left(\alpha^2 - \boldsymbol{\beta}_2'(\boldsymbol{\Delta}_{22} - \boldsymbol{\Delta}_{21}\boldsymbol{\Delta}_{11}^{-1}\boldsymbol{\Delta}_{12})\boldsymbol{\beta}_2\right)^{1/2}$, $\boldsymbol{\beta}^* = \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2'\boldsymbol{\Delta}_{21}\boldsymbol{\Delta}_{11}^{-1}$, $\delta^* = \delta|\boldsymbol{\Delta}_{11}|^{1/(2l)}$, $\boldsymbol{\mu}^* = \boldsymbol{\mu}_1$ and $\boldsymbol{\Delta}^* = |\boldsymbol{\Delta}|^{1/l}\boldsymbol{\Delta}_{11}$.

(b) The conditional distribution of $\boldsymbol{X}_2$ given $\boldsymbol{X}_1 = \boldsymbol{x}_1$ is the k–dimensional generalized hyperbolic distribution $GH_k(\tilde{\lambda}, \tilde{\alpha}, \tilde{\boldsymbol{\beta}}, \tilde{\delta}, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Delta}})$, where $\tilde{\lambda} = \lambda - l/2$, $\tilde{\alpha} = \alpha|\boldsymbol{\Delta}_{11}|^{1/(2k)}$, $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}_2$, $\tilde{\delta} = |\boldsymbol{\Delta}_{11}|^{-1/(2k)} \left(\delta^2 + (\boldsymbol{x}_1 - \boldsymbol{\mu}_1)'\boldsymbol{\Delta}_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1)\right)^{1/2}$, $\tilde{\boldsymbol{\mu}} = \boldsymbol{\mu}_2 + (\boldsymbol{x}_1 - \boldsymbol{\mu}_1)'\boldsymbol{\Delta}_{11}^{-1}\boldsymbol{\Delta}_{12}$ and $\tilde{\boldsymbol{\Delta}} = |\boldsymbol{\Delta}_{11}|^{1/k}(\boldsymbol{\Delta}_{22} - \boldsymbol{\Delta}_{21}\boldsymbol{\Delta}_{11}^{-1}\boldsymbol{\Delta}_{12})$.

(c) Let $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{B}$ be an affine transformation of $\boldsymbol{X}$ and let $||\boldsymbol{A}||$ denote the absolute value of the determinant of $\boldsymbol{A}$. Then, the distribution of $\boldsymbol{Y}$ is the d–dimensional generalized hyperbolic distribution $GH_d(\lambda^+, \alpha^+, \boldsymbol{\beta}^+, \delta^+, \boldsymbol{\mu}^+, \boldsymbol{\Delta}^+)$, where $\lambda^+ = \lambda$, $\alpha^+ = \alpha||\boldsymbol{A}||^{-1/d}$, $\boldsymbol{\beta}^+ = \boldsymbol{\beta}(\boldsymbol{A}^{-1})'$, $\delta^+ = \delta||\boldsymbol{A}||^{1/d}$, $\boldsymbol{\mu}^+ = \boldsymbol{\mu}'\boldsymbol{A} + \boldsymbol{B}$ and $\boldsymbol{\Delta}^+ = ||\boldsymbol{A}||^{-2/d}\boldsymbol{A}'\boldsymbol{\Delta}\boldsymbol{A}$.

Hence, the generalized hyperbolic distribution is closed under margining, conditioning and affine transformations. From the first it follows that even if the returns are modeled jointly, their implied individual distributions still exhibit semiheavy tails. Moreover, the distribution of portfolio returns can easily be deduced by combining property (a) and (c). This is an especially attractive property for assessing portfolio risk.

5.3 The Multivariate Generalized Hyperbolic Stochastic Volatility Model with Realized Covariance

In the following we extend the generalized hyperbolic stochastic volatility model given in equations (5.1) to (5.3) to the multivariate case. Moreover, rather than specifying the mean of the stochastic covariance to be a function of daily squared returns, we make use of the high–frequency data by specifying the mean to depend on past observations of the realized covariance matrix.

Let $\boldsymbol{r}_t$ denote the $d \times 1$ vector of daily returns and $\boldsymbol{\Sigma}_t$ the true, i.e. unobservable, stochastic covariance matrix. The normal variance–mean mixture representation of the multivariate generalized hyperbolic distribution discussed in the previous section provides the basis for our multivariate model. However, as the risk premium effect has been proven to be difficult to identify empirically we follow Andersson (2001) and Barndorff-Nielsen (1997) and exclude it in our multivariate model by assuming $\boldsymbol{\beta} = \boldsymbol{0}$. Our multivariate generalized hyperbolic stochastic volatility

model can then be formulated as follows. First, the returns are assumed to be distributed according to the following multivariate normal distribution

$$\boldsymbol{r}_t \sim N_d(\boldsymbol{\mu}, Z\boldsymbol{\Delta}_t) \tag{5.9}$$

where $\boldsymbol{\Sigma}_t = Z\boldsymbol{\Delta}_t$ and $\boldsymbol{\Delta}_t$ is a $d \times d$ structure matrix that is specified by

$$\boldsymbol{\Delta}_t = \boldsymbol{H}_t(\boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\theta}). \tag{5.10}$$

Moreover,

$$Z \sim GIG(\lambda, \delta, \alpha) \tag{5.11}$$

with $\mathrm{E}[Z] = 1$, which ensures the identifiability of the model parameters such that we do not require that $|\boldsymbol{\Delta}_t| = 1$. Hence,

$$\mathrm{E}[\boldsymbol{\Sigma}_t | \boldsymbol{\mathcal{X}}_{t-1}] = \boldsymbol{H}_t(\boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\theta}).$$

The resulting distribution for the conditional returns is a $GH_d(\lambda, \alpha, \mathbf{0}, \delta, \boldsymbol{\mu}, \boldsymbol{\Delta}_t)$ with mean and covariance given by

$$\begin{aligned} \mathrm{E}[\boldsymbol{r}_t | \boldsymbol{\mathcal{X}}_{t-1}] &= \boldsymbol{\mu} && (5.12) \\ \mathrm{Cov}[\boldsymbol{r}_t | \boldsymbol{\mathcal{X}}_{t-1}] &= \mathrm{E}[Z]\,\boldsymbol{\Delta}_t && (5.13) \\ &= \boldsymbol{H}_t(\boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\theta}). \end{aligned}$$

Obviously, our multivariate model given in equations (5.9) to (5.11) is a straightforward generalization of the univariate case. In particular, it replaces the variance through the covariance matrix (rather than a vector of logarithmic volatilities as is the case in the standard multivariate stochastic volatility models), whose dynamics is modeled via the matrix–variate mean specification $\boldsymbol{H}_t(\boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\theta})$. Alternatively, assuming a constant structure matrix $\boldsymbol{\Delta}$, Barndorff-Nielsen (1997) suggested to model the dynamics of the covariance matrix $Z_t\boldsymbol{\Delta}$ by a time–dependent scalar mean specification for the GIG distributed random variable Z_t, or more precisely by a time–dependent specification of the GIG parameter δ, i.e. $\delta_t = h_t(\boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\theta})$. Such approach, however, implies constant conditional correlations as the full covariance matrix is only driven by a single volatility factor. In view of the often observed correlation breakdown in high volatility periods, especially during downward market movements, such assumption might be too restrictive empirically.

Note that the setup given in equations (5.9) to (5.11) is quite general and allows for various model specifications depending on the particular choice of $\boldsymbol{H}_t(\boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\theta})$. In the following we present two specifications that depend on the realized covariance measure, i.e. $\boldsymbol{\mathcal{X}}_{t-1}$ subsumes only past observations of the realized covariance matrix $\boldsymbol{RC}$. This approach facilitates the construction of a positive definite covariance matrix (as long as the number of assets considered does not exceed the sampling frequency at which the realized covariance matrix is computed), and additionally

allows to exploit the information contained in the high frequency data.[8] Moreover, the use of the full realized covariance matrix immediately results in dynamic conditional correlations.[9]

5.3.1 The Multivariate HAR Model

From the preceding chapters we know, that the dynamics of the individual volatilities can quite adequately be described by the heterogeneous autoregressive (HAR) model. In particular, we have seen that the sum over a few multi–period volatility components allows to reproduce the strong persistence in the volatilities. With the covariances exhibiting similar long–memory patterns, a multivariate version of the HAR model may, thus, be a natural starting point for modeling the mean of the unknown covariance matrix. Of course, our model differs from the existing HAR models as it just determines the mean of the unknown covariance rather than describing the dynamics of the realized (co)variance itself.

To this end we define the k–period realized covariance matrix by

$$\boldsymbol{RC}_{t+1-k:t} = \frac{1}{k}\sum_{j=1}^{k}\boldsymbol{RC}_{t-j}.$$

Then, our multivariate HAR mean specification with daily, weekly and monthly covariance components is given by

$$\boldsymbol{H}_t(\boldsymbol{\mathcal{X}}_{t-1};\boldsymbol{\theta}) = \boldsymbol{A} + \alpha_d\boldsymbol{RC}_{t-1} + \alpha_w\boldsymbol{RC}_{t-5:t-1} + \alpha_m\boldsymbol{RC}_{t-22:t-1}, \tag{5.14}$$

with $\boldsymbol{A}$ a positive definite and symmetric parameter matrix and positive parameters α_d, α_w and α_m. Obviously, for $\alpha_d = \alpha_w = \alpha_m = 0$ we obtain the i.i.d. generalized hyperbolic stochastic volatility model. Restricting these parameters to be scalars implies that the impact of each element of the covariance components on the corresponding element of the expected conditional covariance matrix is the same across all assets. Furthermore, the model does not allow for any cross–sectional inter–dependencies. Such effects are explicitly incorporated in our alternative specification presented in the next section.

Note that a multivariate version of the HAR model was also considered in Bauer and Vorkink (2007), who treat the conditional covariance as being observable via the realized covariance, and, thus, model the realized covariance directly. Moreover, they exclusively focus on the dynamics of the covariance matrix, i.e. leaving the returns unspecified.

[8] So, note that even for a single asset our model specification differs from the NIG stochastic volatility models considered in the literature so far, as those make only use of daily data.

[9] Note that these are not obtainable in the specification suggested by Barndorff-Nielsen (1977). However, the high–frequency information can also be incorporated. For example, the mean of Z_t (or δ_t) may be specified as a function of the realized variance of a stock index, e.g. a univariate HAR–type specification, which may then represent the volatility factor. Alternatively, the factor may be constructed by $\boldsymbol{\psi}'\boldsymbol{RC}_{t-1}\boldsymbol{\psi}$ with $\boldsymbol{\psi}$ denoting an $d \times 1$ parameter vector.

5.3.2 The Quadratic Model

Our second specification is closely related to the multivariate stochastic volatility model of Philipov and Glickman (2006). Let $\boldsymbol{B} = (\boldsymbol{B}^{1/2})(\boldsymbol{B}^{1/2})'$ be a positive definite and symmetric parameter matrix that is decomposed by the Cholesky decomposition, and let $\boldsymbol{A}$ and $\boldsymbol{RC}_{t-5:t-1}$ be defined as before. We then specify the following quadratic model for the mean of the covariance matrix

$$\boldsymbol{H}_t(\boldsymbol{\mathcal{X}}_{t-1};\boldsymbol{\theta}) = \boldsymbol{A} + (\boldsymbol{B}^{1/2})\boldsymbol{RC}_{t-5:t-1}(\boldsymbol{B}^{1/2})', \tag{5.15}$$

ensuring positive definiteness. Again, if $\boldsymbol{B}$ is zero, the model reduces to the i.i.d. multivariate generalized hyperbolic stochastic volatility model. For $\boldsymbol{B}$ nonzero and non–diagonal, this model has an especially attractive property. In particular, the individual conditional (co)variances are then allowed to depend not only on the history of their own realized (co)variance but also on the realized covariances and variances of other assets. Thus, the model introduces intertemporal cross–sectional effects, such as the often documented contagion among asset returns, by letting the conditional (co)variances depend on their past as measured by the lagged monthly realized covariance component. As pointed out by Philipov and Glickman (2006), this feature is not present in other multivariate stochastic volatility models. Their model specification is in fact very similar to ours. However, they assume an inverse Wishart distribution for the stochastic covariance matrix with mean depending on the preceding lag of the latent stochastic covariance rather than the realized covariance or realized covariance component. As a consequence, the estimation of their model is nontrivial leading them to develop a suitable MCMC algorithm. In contrast, when using the realized covariance matrix the model can straightforwardly be estimated via maximum–likelihood. Note that the use of a multiperiod rather than the daily realized covariance component allows us to incorporate more past information into the mean specification.

Note that the introduction of the intertemporal cross–sectional inter–dependencies in the elements of the covariance matrix comes of course at the cost of parameter parsimony. In particular, the number of parameters in the model is $d(d+1)$. The multivariate HAR model instead is less subject to this so–called "curse of dimensionality" and may therefore be preferable in practical applications involving large scale portfolios.

5.4 Empirical Application

We now turn to the application of our model to the three daily S&P500 component stocks discussed in Section 2.2, i.e. INTC, MSFT and PFE. Both model specifications are fitted using the daily return series and the realized covariance measure constructed from the 30 minute intraday returns. Following Andersson (2001), Barndorff-Nielsen (1997) and Jensen and Lunde (2001) we focus in the sequel on the case that $\lambda = -\frac{1}{2}$. Hence, the mixing distribution is now given by the inverse

Gaussian distribution, which is the generalized inverse Gaussian distribution with $\lambda = -\frac{1}{2}$. Consequently, we fit a multivariate normal inverse Gaussian stochastic volatility model.

Moreover, in order to assess the relevance of the stochastic volatility specification we also estimate a Gaussian model. In particular, we assume that the returns are conditionally multivariate normal distributed with deterministic conditional variance given by the respective specifications of $\boldsymbol{H}_t(\boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\theta})$. The model can thus be interpreted as a multivariate ARCH–type model based, however, on high–frequency returns.

Table 5.1 presents the estimation results for the multivariate HAR models. Note that the reported estimates are based on a restricted multivariate HAR specification. In particular, the estimation of the model with the HAR mean specification as given equation (5.14) yielded nearly zero and insignificant coefficient of the lagged daily realized covariance component. We have therefore excluded the daily component and have re–estimated the models. Obviously, with the exception of the

Table 5.1: Estimation Results of the Multivariate HAR Models

	NIG–SV–HAR			Gaussian–HAR		
		$\hat{\boldsymbol{A}}$			$\hat{\boldsymbol{A}}$	
INTC	1.6075			1.4945		
	(0.0771)			(0.0625)		
MSFT	0.5137	1.2016		0.4764	1.1641	
	(0.0330)	(0.0431)		(0.0289)	(0.0363)	
PFE	0.1087	0.1308	1.0519	0.0726	0.0951	0.9953
	(0.0287)	(0.0244)	(0.0450)	(0.0252)	(0.0221)	(0.0326)
α_w		0.0865			0.1000	
		(0.0227)			(0.0197)	
α_m		0.0767			0.0896	
		(0.0246)			(0.0212)	
$\boldsymbol{\mu}'$	-0.1081	-0.0571	-0.1418	-0.0789	-0.0291	-0.1267
	(0.0551)	(0.0405)	(0.0363)	(0.0692)	(0.0498)	(0.0401)
α		1.3630				
		(0.0903)				
Log–L.		6,901.9			7,018.2	

Presented are the maximum–likelihood estimates of the normal inverse Gaussian stochastic volatility model with the multivariate HAR mean specification (NIG–SV–HAR) and of the Gaussian model with the multivariate HAR specification for the covariance (GAUSSIAN–HAR). The numbers in parentheses report the corresponding standard errors.

conditional mean of MSFT (i.e. μ_{MSFT}) all parameters of these restricted models are highly significant. Generally, the estimated coefficients of the volatility components differ from those reported in the literature (and in the preceding two chapters) for the univariate HAR model. This may be expected as in the multivariate model these coefficients do not only measure the impact of the (multiperiod) lagged realized variances but also of the (multiperiod) lagged realized covariances on the respective elements of the covariance matrix. Furthermore, in contrast to the existing HAR models, that describe the dynamics of the realized (co)variance, our specification determines the mean of the unknown covariance. Still, similarly to the estimation results of the univariate HAR models presented in Chapter 3, the HAR coefficients of the normal inverse Gaussian specification are smaller than those of the Gaussian model. Importantly, the log likelihood values reported in the last row of Table 5.1 clearly show the superior fit of our normal inverse Gaussian stochastic volatility model relative to the Gaussian ARCH–type model.[10]

Table 5.2 presents the maximum–likelihood estimates of the quadratic models. Also in this case, the normal inverse Gaussian stochastic volatility model strongly outperforms the Gaussian ARCH–like model. The results, thus, clearly indicate that a stochastic specification of the covariance matrix is empirically very important.

Figure 5.1 depicts the estimated variances and correlations of the multivariate HAR and of the quadratic normal inverse Gaussian stochastic volatility models. Obviously, the series are much smoother than the realized (co)variances (see Figure 2.9). This finding is not surprising as the model–implied volatilities are averages over multiple lagged realized covariances. Moreover, based on the stochastic specification of the volatility, the model may also incorporate part of the noisiness inherent in the realized covariance measure. The Figure also illustrates that the variances and correlations implied by the quadratic model are generally more volatile and move together more closely, which results from the incorporation of intertemporal cross–sectional dependencies.

Note that none of the two specifications is formally a long–memory model. Figures 5.2 and 5.3, however, illustrate for both models that the shapes of the sample autocorrelation functions of the estimated (co)variances are quite similar to those observed for the realized covariance measure (see Figure 2.8). However, they tend to overestimate the autocorrelation especially for the lags of up to one month, which is due to the averaging over the daily realized covariances. These patterns are more pronounced in the quadratic model. Allowing for contagion–like effects also seems to lead to more distinct swings, i.e. comovements, in the autocorrelation functions of all (co)variance series.

[10] Note that the log likelihood values can straightforwardly be compared with each other, as the Gaussian model is nested in the stochastic volatility model.

Table 5.2: Estimation Results of the Quadratic Models

	NIG–SV–Q			Gaussian–Q		
	$\hat{\boldsymbol{A}}$			$\hat{\boldsymbol{A}}$		
INTC	1.0853			0.9630		
	(0.1167)			(0.1029)		
MSFT	-0.1695	0.7615		0.0584	0.8806	
	(0.0699)	(0.0800)		(0.0753)	(0.0590)	
PFE	0.3090	0.2199	1.1364	0.3027	0.1367	1.1002
	(0.0515)	(0.0423)	(0.0501)	(0.0356)	(0.0221)	(0.0418)
	$\hat{\boldsymbol{B}}$			$\hat{\boldsymbol{B}}$		
INTC	0.5758			0.5894		
	(0.0315)			(0.0288)		
MSFT	0.3671	0.2696		0.1868	0.4779	
	(0.0492)	(0.0910)		(0.0370)	(0.0381)	
PFE	-0.1561	0.0934	0.3316	-0.1685	0.0970	0.3420
	(0.0313)	(0.0717)	(0.0340)	(0.0284)	(0.0459)	(0.0236)
$\boldsymbol{\mu}'$	-0.0571	-0.0283	-0.1124	0.0245	0.0333	-0.0858
	(0.0546)	(0.0380)	(0.0367)	(0.0574)	(0.0437)	(0.0397)
α		1.3527				
		(0.0886)				
Log–L.		6,832.4			6,947.4	

Presented are the maximum–likelihood estimates of the normal inverse Gaussian stochastic volatility model with the quadratic mean specification (NIG–SV–Q) and of the Gaussian model with the quadratic specification for the covariance (GAUSSIAN–Q). The numbers in parentheses report the corresponding standard errors.

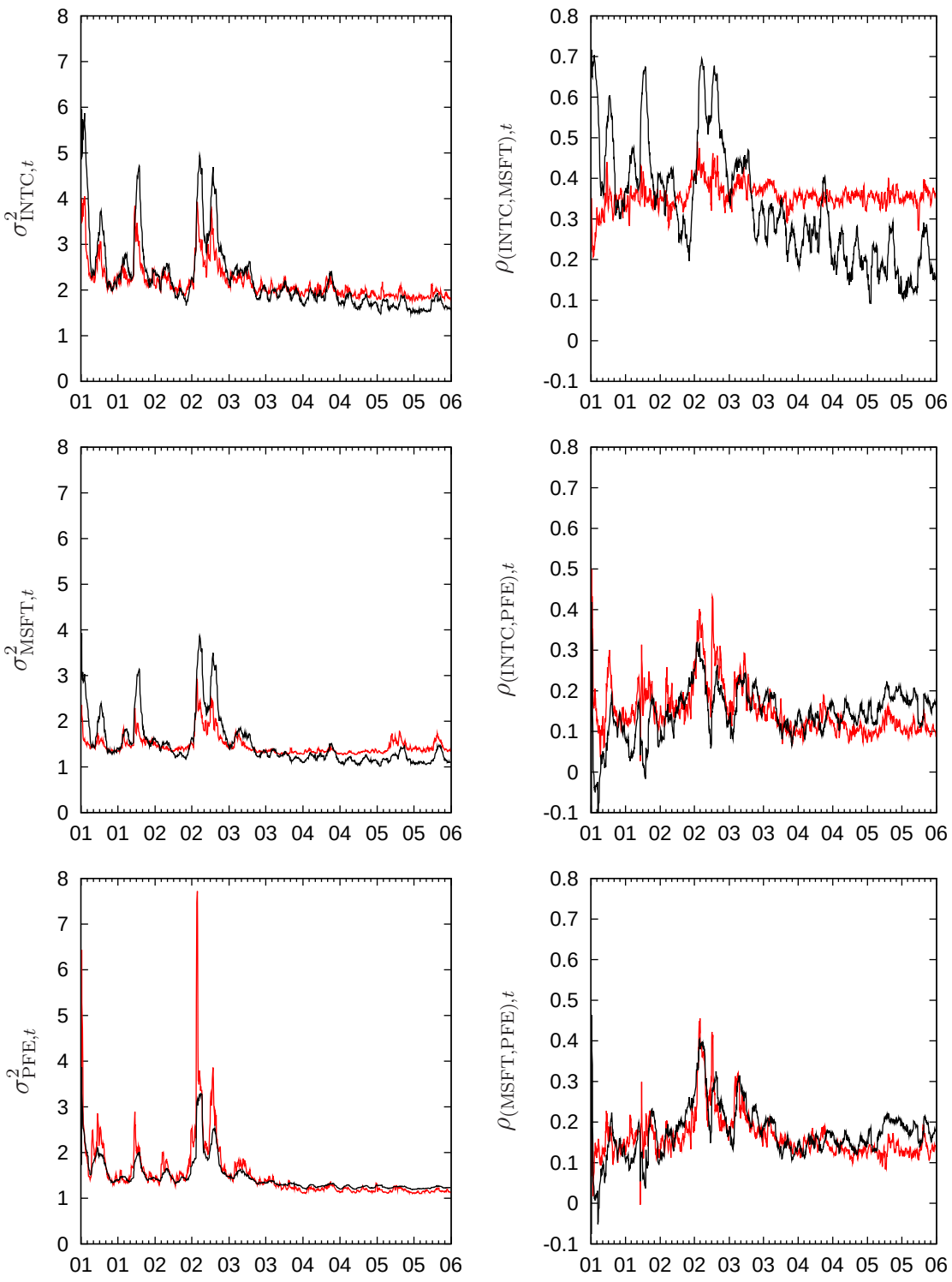

Figure 5.1: Depicted are the estimated variances and correlations of the NIG–SV–HAR model (red line) and of the NIG–SV–Q model (black line). The left panels present the estimated variances, whereas the right panels present the estimated correlations.

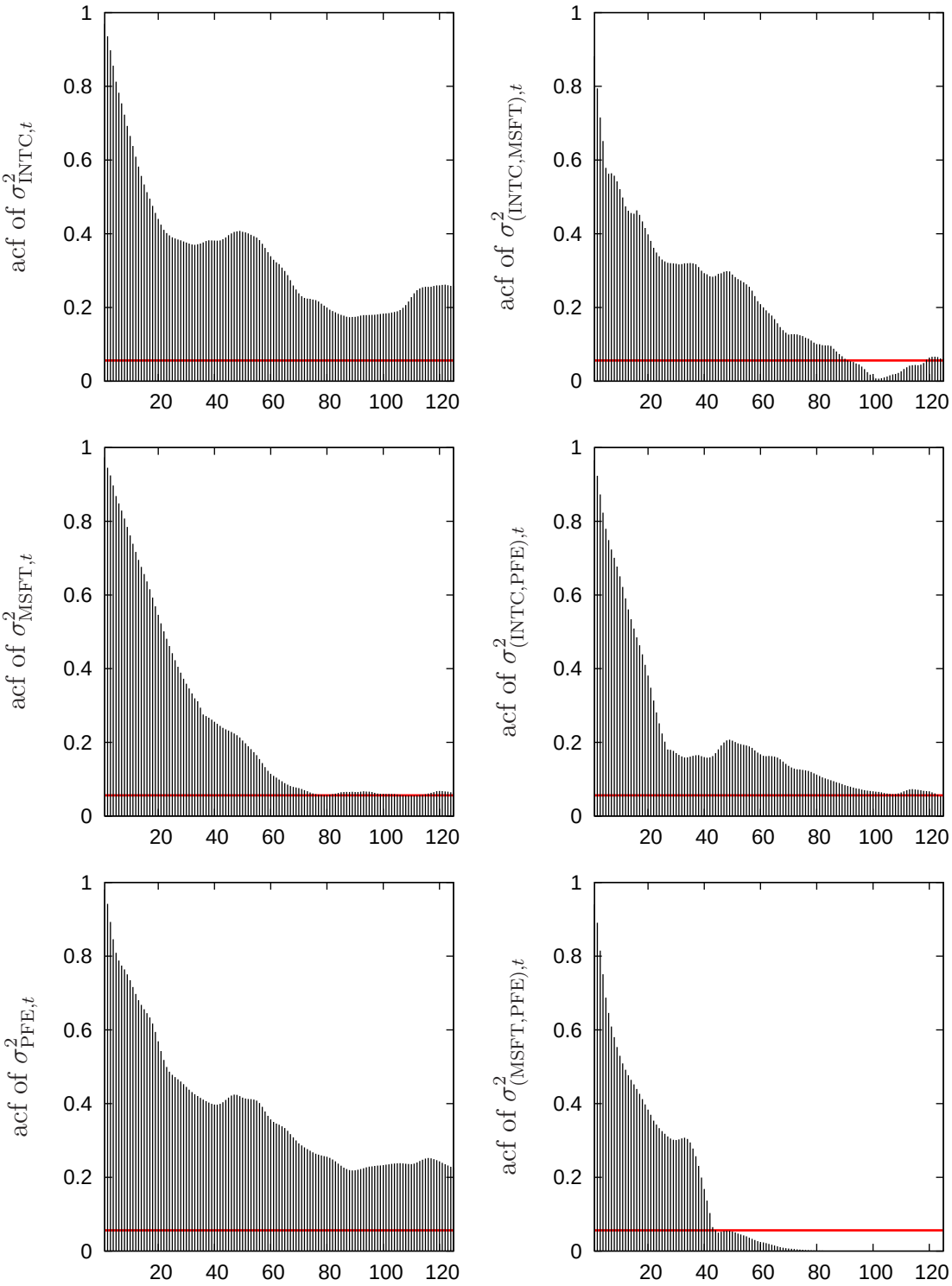

Figure 5.2: Sample autocorrelations of the estimated variances and covariances based on the HAR model.

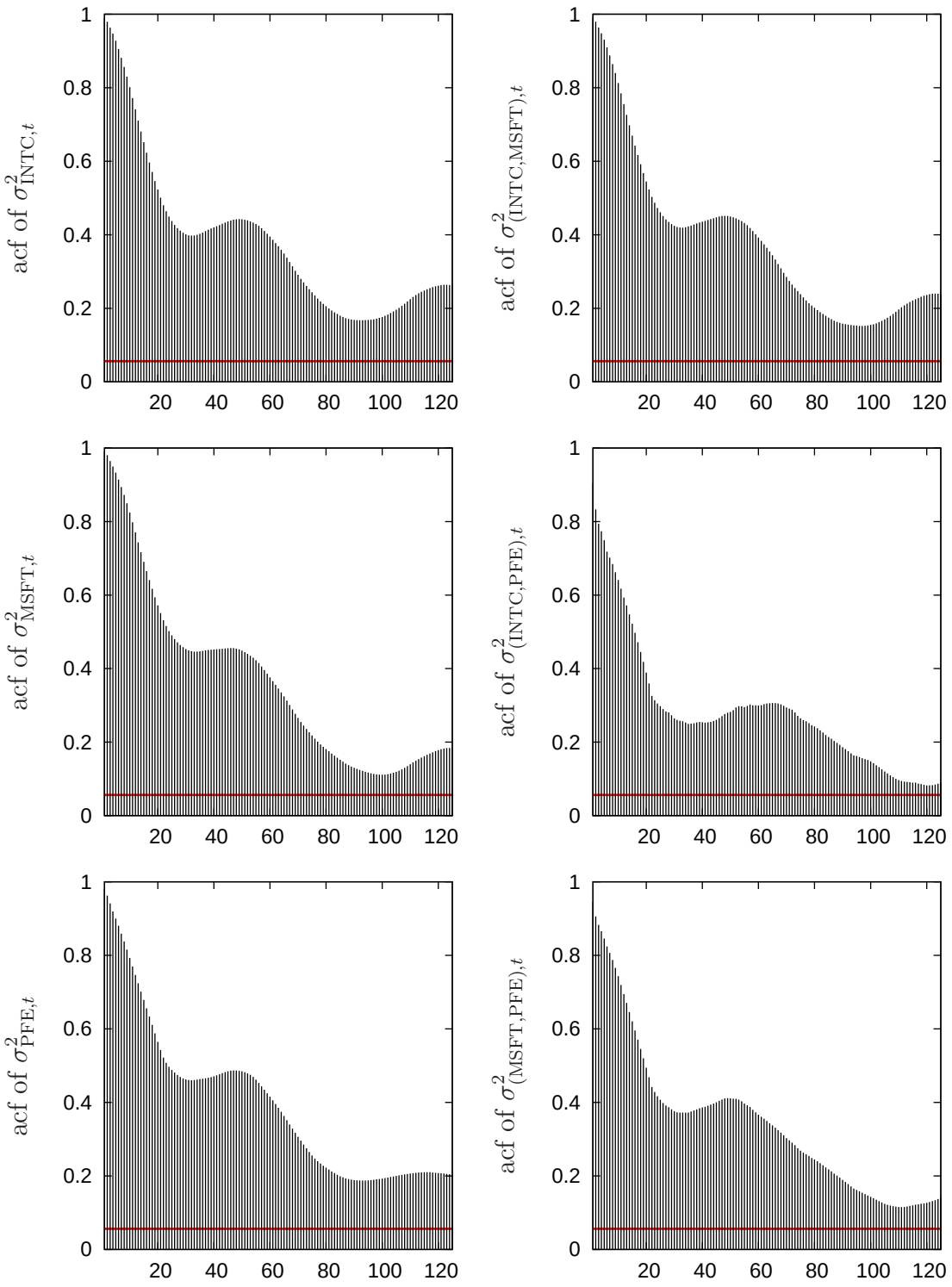

Figure 5.3: Sample autocorrelations of the estimated variances and covariances based on the quadratic model.

5.5 Conclusion

This Chapter developed a multivariate generalization of the normal inverse Gaussian stochastic volatility model of Barndorff-Nielsen (1997). The modeling framework is especially appealing as it allows to incorporate the information contained in the various volatility measures, such as the realized covariance, while acknowledging the fact, that these series are noisy measures of the unknown covariance. Hence, specifying a stochastic covariance matrix whose mean is given as a deterministic function of the variation measures combines stochastic volatility and GARCH–type features. In fact we have shown for the univariate case, that the generalized hyperbolic stochastic volatility model and the generalized hyperbolic GARCH model are equivalent, if the risk premia effect is excluded. This is an interesting result as most of the popular GARCH models can thus also be interpreted as Gaussian stochastic volatility models with mixing stochastic variance whose mean is given by a GARCH–type specification. Similar arguments should also apply to the multivariate model.

The multivariate generalized hyperbolic stochastic volatility model in fact is a straightforward generalization of the univariate case as the variance is simply replaced by the covariance matrix. Thus, the full covariance matrix is modeled directly via a matrix–variate mean specification, which allows the straightforward implementation of dynamic conditional correlations, e.g. by using the realized covariance measure. Moreover, our general model setup is quite flexible and allows for various choices of the mean function. Thus depending on the application at hand the mean function may be chosen correspondingly. E.g. if large scale portfolios are of main interest one may favor a less flexible but more parsimonious mean specification, such as our multivariate HAR model. In contrast, for a smaller number of assets or for assets that are well–known to be strongly correlated, the quadratic model may be suitable as it explicitly allows for intertemporal cross–sectional dependencies. Importantly, comparing our model specifications with a multivariate realized covariance based Gaussian ARCH–type model, our empirical application also reveals that the inclusion of a stochastic covariance matrix is indeed essential.

A more thorough analysis of the performance of our models within a portfolio optimization application, e.g. the performance of global minimum–variance portfolios, will be the subject of future research. Moreover, as the distribution of portfolio returns can easily be deduced in our models, a forecasting and portfolio risk management analysis may be of additional interest.

6 Conclusion

The recent availability of the high–frequency–based realized variation measures, has lead to the development of new financial volatility models that have been shown to generally outperform the traditional GARCH or stochastic volatility models in terms of model adequacy and forecasting ability. Motivated by these findings, we also make use here of the realized variation measures. However, rather than focusing solely on the realized volatility as a measure of total price variation, we consider additional realized measures constructed from the high–frequency data in order to assess new and important aspects of the dynamics of stock returns.

In particular, we show that the realized volatility measure exhibits important volatility clustering, which is also present in the innovations of the most commonly used realized volatility models. Accounting explicitly for this time–variation in the volatility of volatility by specifying a GARCH model results in further improvements in the in–sample as well as the out–of–sample performance of the heterogeneous autoregressive realized volatility model. Within this modeling framework we also find that a more flexible, i.e. a skewed and leptokurtic, innovation distribution relative to the Gaussian is more adequate. Moreover, formulating the models in terms of the logarithm of the realized variance is more robust to the observed heteroscedasticity and non–Gaussianity and may be preferred in forecasting applications.

Utilizing that the sum over the product of absolute current and lagged high–frequency returns provides a measure that is robust to jumps in the price process, we decompose the total price variation into the variation coming from the continuous sample path evolvement and that coming from the jumps. This allows us to assess separately the different dynamic and distributional properties of the two volatility components and their interrelationship with daily returns within a coherent framework. Importantly, developing a simultaneous equation model we show that the often observed lagged leverage effect primarily acts through the continuous volatility component (as measured by the Bipower variation) and that there exists a similar mechanism in the contemporaneous leverage effect. This is an interesting result, as it also supports the type of leverage specification commonly used in the continuous–time stochastic volatility models. Moreover, the stunning accuracy of our model along with the availability of its likelihood function and analytic derivatives makes it an ideal candidate as an auxiliary model for the estimation of continuous–time stochastic volatility models using indirect inference methods. More precisely, our model is highly informative about the empirical properties that should be captured by these models and, thus, helps in assessing and developing adequate continuous–time stochastic volatility models using the high–frequency in-

formation inherent in the realized variation measures. In fact, employing this idea, we show in Bollerslev et al. (2006a) that the most commonly used affine and logarithmic jump diffusion models miss important features of the data, such as the volatility persistence and the leverage effect.

Rather than modeling solely the individual dynamics of stock returns, we also exploit the information contained in the realized covariance measure for modeling jointly the cross–sectional dynamics of multiple assets. Extending the generalized hyperbolic stochastic volatility model to the multivariate case, our approach differs from the existing realized covariance literature, as we, firstly, model jointly the covariance and the returns, and, secondly, treat the covariance as being unobservable. In particular, we account for the fact, that the realized covariance is an unbiased but a noisy measure of the true covariance. Moreover, our modeling strategy shows, that the use of the realized covariance measure in multivariate modeling has several advantages over the standard multivariate stochastic volatility models. In particular, the realized covariance matrix can be directly used to model the full covariance matrix and as such allows the straightforward implementation of dynamic conditional correlations. In addition, the positive–definiteness of the realized covariance matrix allows to consider a generally larger number of assets as would be the case with daily data. And, last but not least, while our model is still a stochastic volatility model, its maximum–likelihood estimation becomes straightforward, as is illustrated by our empirical application. The generality of our multivariate model also allows to incorporate various types of dependencies in the covariance matrix. This is exemplified by our particular specifications for the mean function of the covariance matrix. However, just like any other specification of a multivariate stochastic volatility or GARCH–type model, our model, or more precisely, the specific choice of the mean function, is subject to the trade–off between model parsimony and flexibility.

Bibliography

Agiakloglou, C., Newbold, P., and Wohar, M. (1993), "Bias in an Estimator of the Fractional Difference Parameter," *Journal of Time Series Analysis*, 14, 235–246.

Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2005), "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise," *Review of Financial Studies*, 18, 351–416.

Andersen, T. G., Benzoni, L., and Lund, J. (2002), "An Empirical Investigation of Continuous-Time Equity Return Models," *Journal of Finance*, 57, 1239–1284.

Andersen, T. G. and Bollerslev, T. (1997), "Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns," *The Journal of Finance*, 52, 975–1005.

— (1998), "Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts," *International Economic Review*, 39, 885–905.

Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2007), "Roughing it Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility," *Review of Economics and Statistics*, forthcoming.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001a), "The Distribution of Realized Stock Return Volatility," *Journal of Financial Economics*, 61, 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001b), "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42–55.

— (2003), "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

Andersen, T. G., Bollerslev, T., and Meddahi, N. (2005), "Correcting the Errors: Volatility Forecast Evaluation Using High-Frequency Data and Realized Volatilities," *Econometrica*, 73, 279–296.

Andersson, J. (2001), "On the Normal Inverse Gaussian Stochastic Volatility Model," *Journal of Business & Economic Statistics*, 19, 44–54.

Asai, M. and McAleer, M. (2005), "Dynamic Asymmetric Leverage in Stochastic Volatility Models," *Econometric Reviews*, 24, 317–332.

— (2006), "Asymmetric Multivariate Stochastic Volatility," *Econometric Reviews*, 25, 453–473.

Asai, M., McAleer, M., and Yu, J. (2006), "Multivariate Stochastic Volatility: A Review," *Econometric Reviews*, 25, 145–175.

Bai, X., Russell, J. R., and Tiao, G. C. (2003), "Kurtosis of GARCH and Stochastic Volatility Models with Non-Normal Innovations," *Journal of Econometrics*, 114, 349–360.

Baillie, R. T., Chung, C.-F., and Tieslau, M. A. (1996), "Analysing Inflation by the Fractionally Integrated ARFIMA-GARCH Model," *Journal of Applied Econometrics*, 11, 23–40.

Bandi, F. M. and Russell, J. R. (2005a), "Microstructure Noise, Realized Volatility, and Optimal Sampling," Working Paper, The University of Chicago.

— (2005b), "Realized Covariation, Realized Beta, and Microstructure Noise," Working Paper, The University of Chicago.

Banerjee, A. and Urga, G. (2005), "Modelling Structural Breaks, Long Memory and Stock Market Volatility: An Overview," *Journal of Econometrics*, 129, 1–2.

Barndorff-Nielsen, O. E. (1977), "Exponentially Decreasing Distributions for the Logarithm of Particle Size," *Proceedings of the Royal Society of London. Series A*, 353, 401–419.

— (1995), "Normal Inverse Gaussian Process and the Modelling of Stock Returns," Research report 300, Department of Theoretical Statistics, University of Aarhus.

— (1997), "Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling," *Scandinavian Journal of Statistics*, 24, 1–13.

— (1998), "Processes of Normal Inverse Gaussian Type," *Finance and Stochastics*, 2, 41–68.

Barndorff-Nielsen, O. E. and Blæsild, P. (1981), "Hyperbolic Distributions and Ramifications: Contributions to Theory and Application," in *Statistical Distributions in Scientific Work*, eds. Taillie, C., Patil, G. P., and Baldessari, B. A., Dordrecht: Reidel, vol. 4, pp. 19–44.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2006a), "Designing Realised Kernels to Measure the ex-post Variation of Equity Prices in the Presence of Noise," Working Paper, University of Oxford.

Barndorff-Nielsen, O. E. and Shephard, N. (2001), "Modelling by Lévy Processess for Financial Econometrics," in *Lévy Processes – Theory and Applications*, eds. Barndorff-Nielsen, O. E., Mikosch, T., and Resnick, S. I., Basel: Birkhäuser, pp. 283–318.

— (2002a), "Econometric Analysis of Realised Volatility and its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society, Series B*, 64, 253–280.

— (2002b), "Estimating Quadratic Variation Using Realized Variance," *Journal of Applied Econometrics*, 17, 457–477.

— (2003), "Realised Power Variation and Stochastic Volatility Models," *Bernoulli*, 9, 243–265.

— (2004a), "Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics," *Econometrica*, 72, 885–925.

— (2004b), "Power and Bipower Variation with Stochastic Volatility and Jumps," *Journal of Financial Econometrics*, 2, 1–37.

— (2005), "How Accurate is the Asymptotic Approximation to the Distribution of Realised Variance?" in *Identification and Inference for Econometric Models. A Festschrift for Tom Rothenberg*, eds. Andrews, D., Powell, J., Ruud, P., and Stock, J., Cambridge: Cambridge University Press, pp. 306–331.

— (2006), "Econometrics of Testing for Jumps in Financial Economics using Bipower Variation," *Journal of Financial Econometrics*, 4, 1–30.

Barndorff-Nielsen, O. E., Shephard, N., and Winkel, M. (2006b), "Limit theorems for multipower variation in the presence of jumps," *Stochastic Processes and their Applications*, 116, 796–806.

Bates, D. S. (2000), "Post-'87 Crash Fears in the S&P 500 Futures Option Market," *Journal of Econometrics*, 94, 181–238.

Bauer, G. H. and Vorkink, K. (2007), "Multivariate Realized Stock Market Volatility," Working Paper, Bank of Canada.

Bekaert, G. and Wu, G. (2000), "Asymmetric Volatility and Risk in Equity Markets," *Review of Financial Studies*, 13, 1–42.

Beran, J. (1994), *Statistics for Long-Memory Processes*, Chapman & Hall.

Blæsild, P. and Jensen, J. L. (1981), "Multivariate Distributions of Hyperbolic Type," in *Statistical Distributions in Scientific Work*, eds. Taillie, C., Patil, G. P., and Baldessari, B. A., Dordrecht: Reidel, vol. 4, pp. 45–66.

Bollen, B. and Inder, B. (2002), "Estimating Daily Volatility in Financial Markets Utilizing Intraday Data," *Journal of Empirical Finance*, 9, 551–562.

Bollerslev, T. (1987), "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," *Review of Economics and Statistics*, 69, 542–547.

Bollerslev, T., Gallant, R., Pigorsch, C., Pigorsch, U., and Tauchen, G. (2006a), "Statistical Assessment of Models for Very High Frequency Financial Price Dynamics," Working Paper, Duke University.

Bollerslev, T., Kretschmer, U., Pigorsch, C., and Tauchen, G. (2007), "A Discrete-Time Model for Daily S&P500 Returns and Realized Variations: Jumps and Leverage Effects," *Journal of Econometrics*, forthcoming.

Bollerslev, T., Litvinova, J., and Tauchen, G. (2006b), "Leverage and Volatility Feedback Effects in High-Frequency Data," *Journal of Financial Econometrics*, 4, 353–384.

Bollerslev, T. and Zhou, H. (2002), "Estimating Stochastic Volatility Diffusions Using Conditional Moments of Integrated Volatility," *Journal of Econometrics*, 109, 33–65.

Calvet, L. and Fisher, A. (2002), "Multifractality in Asset Returns: Theory and Evidence," *Review of Economics and Statistics*, 84, 381–406.

— (2004), "How to Forecast Long-Run Volatility: Regime Switching and the Estimation of Multifractal Processes," *Journal of Financial Econometrics*, 2, 49–83.

Campbell, J. Y. and Hentschel, L. (1992), "No News is Good News: An Asymmetric Model of Changing Volatility in Stock Returns," *Journal of Financial Economics*, 31, 281–331.

Chan, D., Kohn, R., and Kirby, C. (2006), "Multivariate Stochastic Volatility with Correlated Errors," *Econometric Reviews*, 25, 245–274.

Chernov, M., Gallant, A. R., Ghysels, E., and Tauchen, G. (2003), "Alternative Models for Stock Price Dynamics," *Journal of Econometrics*, 116, 225–257.

Chib, S., Nardari, F., and Shephard, N. (2006), "Analysis of High-Dimensional Multivariate Stochastic Volatility Models," *Journal of Econometrics*, 134, 341–371.

Chiriac, R. and Voev, V. (2007), "Long Memory Modelling of Realized Covariance Matrices," Working Paper, University of Konstanz.

Clark, P. K. (1973), "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices," *Econometrica*, 41, 135–155.

Clements, M. P. and Hendry, D. F. (1998), *Forecasting Economic Time Series*, Cambridge: Cambridge University Press.

Comte, F. and Renault, E. (1998), "Long-Memory in Continuous-Time Stochastic Volatility Models," *Mathematical Finance*, 8, 291–323.

Corradi, V., Distaso, W., and Swanson, N. R. (2005), "Predictive Density Estimators for Daily Volatility Based on the Use of Realized Measures," Working Paper, Queen Mary, University of London and Rutgers University.

— (2006), "Predictive Inference for Integrated Volatility," Working Paper, Queen Mary, University of London and Rutgers University.

Corsi, F. (2004), "A Simple Long Memory Model of Realized Volatility," Working Paper, University of Southern Switzerland.

Corsi, F., Mittnik, S., Pigorsch, C., and Pigorsch, U. (2007), "The Volatility of Realized Volatility," *Econometric Reviews*, forthcoming.

Corsi, F., Zumbach, G., Müller, U. A., and Dacorogna, M. (2001), "Consistent High-Precision Volatility from High-Frequency Data," *Economic Notes*, 30, 183–204.

Da, Z. and Schaumburg, E. (2006), "The Factor Structure of Realized Volatility and its Implications for Option Pricing," Working Paper, University of Notre Dame.

Danielsson, J. (1994), "Stochastic Volatility in Asset Prices: Estimation with Simulated Maximum Likelihood," *Journal of Econometrics*, 64, 375–400.

Davidson, J. and Teräsvirta, T. (2002), "Long Memory and Nonlinear Time Series," *Journal of Econometrics*, 110, 105–112.

de Pooter, M., Martens, M., and van Dijk, D. (2006), "Predicting the Daily Covariance Matrix for S&P 100 Stocks Using Intraday Data - But Which Frequency to Use?" *Econometric Reviews*, forthcoming.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998), "Evaluating Density Forecasts with Applications to Financial Risk Management," *International Economic Review*, 39, 863–883.

Diebold, F. X. and Inoue, A. (2001), "Long Memory and Regime Switching," *Journal of Econometrics*, 105, 131–159.

Ding, Z. and Granger, C. W. J. (1996), "Modeling Volatility Persistence of Speculative Returns: A New Approach," *Journal of Econometrics*, 73, 185–215.

Ding, Z., Granger, C. W. J., and Engle, R. F. (1993), "A Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance*, 1, 83–106.

Doornik, J. A. and Ooms, M. M. (2003), "Computational Aspects of Maximum Likelihood Estimation of Autoregressive Fractionally Integrated Moving Average Models," *Computational Statistics and Data Analysis*, 42, 333–348.

Eberlein, E. and Keller, U. (1995), "Hyperbolic Distributions in Finance," *Bernoulli*, 1, 281–299.

Eberlein, E., Keller, U., and Prause, K. (1998), "New Insights into Smile, Mispricing, and Value at Risk: The Hyperbolic Model," *Journal of Business*, 71, 371–405.

Eberlein, E. and von Hammerstein, E. A. (2004), "Generalized Hyperbolic and Inverse Gaussian Distributions: Limiting Cases and Approximation of Processes," in *Seminar on Stochastic Analysis, Random Fields and Applications IV*, eds. Dalang, R., Dozzi, M., and Russo, F., Birkhäuser, pp. 221–264.

Engle, R. F. and Lee, G. (1999), "A Permanent and Transitory Component Model of Stock Return Volatility," *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W. J. Granger*, 475–497.

Engle, R. F., Lilien, D. M., and Robbins, R. P. (1987), "Estimating Time Varying Risk Premia in the Term Structure," *Journal of Business & Economic Statistics*, 9, 345–359.

Engle, R. F. and Ng, V. K. (1993), "Measuring and Testing the Impact of News on Volatility," *Journal of Finance*, 48, 1749–1778.

Eraker, B. (2004), "Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices," *Journal of Finance*, 59, 1367–1403.

Eraker, B., Johannes, M., and Polson, N. (2003), "The Impact of Jumps in Volatility and Returns," *Journal of Finance*, 58, 1269–1300.

Forsberg, L. and Bollerslev, T. (2002), "Bridging the Gap Between the Distribution of Realized (ECU) Volatility and ARCH Modeling (of the Euro): The GARCH-NIG Model," *Journal of Applied Econometrics*, 17, 535–548.

French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987), "Expected Stock Returns and Volatility," *Journal of Financial Economics*, 19, 3–29.

Gallant, A. R. and McCulloch, R. E. (2005), "On the Determination of General Scientific Models with Application to Asset Pricing," Working Paper, Duke University.

Gallant, A. R. and Nychka, D. W. (1987), "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica*, 55, 363–390.

Gallant, A. R. and Tauchen, G. (1989), "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications," *Econometrica*, 57, 1091–1120.

— (1996), "Which Moment to Match?" *Econometric Theory*, 12, 657–681.

Geweke, J. and Porter-Hudak, S. (1983), "The Estimation and Application of Long Memory Time Series Models," *Journal of Time Series Analysis*, 4, 221–238.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2005), "There is a Risk-Return Trade-Off After All," *Journal of Financial Economics*, 76, 509–548.

Giot, P. and Laurent, S. (2004), "Modelling Daily Value-at-Risk using Realized Volatility and ARCH Type Models," *Journal of Empirical Finance*, 11, 379–398.

Gonccalves, S. and Meddahi, N. (2005), "Bootstrapping Realized Volatility," Working Paper, University of Montreal.

Gourieroux, C. (2006), "Continuous Time Wishart Process for Stochastic Risk," *Econometric Reviews*, 25, 177–217.

Gourieroux, C. and Jasiak, J. (2001), *Financial Econometrics: Problems, Models, and Methods*, Princeton: Princeton University Press.

Granger, C. W. J. (1980), "Long Memory Relationships and the Aggregation of Dynamic Models," *Journal of Econometrics*, 14, 227–238.

Granger, C. W. J. and Hyung, N. (2004), "Occasional Structural Breaks and Long Memory with an Application to the S&P 500 Absolute Stock Returns," *Journal of Empirical Finance*, 11, 399–421.

Granger, C. W. J. and Joyeux, R. (1980), "An Introduction to Long-Range Time Series Models and Fractional Differencing," *Journal of Time Series Analysis*, 1, 15–30.

Granger, C. W. J., Spear, S., and Ding, Z. (2000), "Stylized Facts on the Temporal and Distributional Properties of Daily Data from Speculative Markets," in *Statistics and Finance: An Interface*, eds. Chan, W. S., Lin, W. K., and Tong, H., London: Imperial College Press, pp. 97–120.

Granger, C. W. J. and Teräsvirta, T. (1999), "A Simple Nonlinear Time Series Model with Misleading Linear Properties," *Economics Letters*, 62, 161–165.

Hansen, P. R. and Lunde, A. (2006), "Consistent Ranking of Volatility Models," *Journal of Econometrics*, 131, 97–121.

Harvey, A. C., Ruiz, E., and Shephard, N. (1994), "Multivariate Stochastic Variance Models," *Review of Economic Studies*, 61, 247–264.

Harvey, A. C. and Spephard, N. (1996), "Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns," *Journal of Business & Economic Statistics*, 14, 429–434.

Hayashi, T. and Yoshida, N. (2005), "On Covariance Estimation of non-synchronously Observed Diffusion Processes," *Bernoulli*, 11, 359–379.

Hosking, J. R. M. (1981), "Fractional Differencing," *Biometrika*, 68, 165–176.

Hsieh, D. A. (1991), "Chaos and Nonlinear Dynamics: Application to Financial Markets," *Journal of Finance*, 46, 1839–1877.

Huang, X. and Tauchen, G. (2005), "The Relative Contribution of Jumps to Total Price Variance," *Journal of Financial Econometrics*, 3, 456–499.

Jacquier, E., Polson, N. G., and Rossi, P. E. (1994), "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business & Economic Statistics*, 12, 281–300.

Jensen, M. B. and Lunde, A. (2001), "The NIG-S&ARCH Model: A Fat-tailed, Stochastic, and Autoregressive Conditional Heteroskedastic Volatility Model," *Econometrics Journal*, 4, 319–342.

Koopman, S. and Uspensky, E. H. (2002), "Stochastic Volatility in Mean Model: Empirical Evidence from International Stock Markets," *Journal of Applied Econometrics*, 17, 667–689.

Koopman, S. J., Jungbacker, B., and Hol, E. (2005), "Forecasting Daily Variability of the S&P 100 Stock Index Using Historical, Realised and Implied Volatility Measurements," *Journal of Empirical Finance*, 12, 445–475.

Küchler, U., Neumann, K., Sørensen, M., and Streller, A. (1999), "Stock Returns and Hyperbolic Distributions," *Mathematical and Computer Modelling*, 29, 1–15.

Lundblad, C. (2004), "The Risk Return Tradeoff in the Long-Run: 1836-2003," Working Paper, University of North Carolina at Chapel Hill.

Lux, T. and Marchesi, M. (1999), "Scaling and Criticality in a Stochastic Multi-Agent Model of a Financial Market," *Nature*, 397, 498–500.

Lütkepohl, H. (1993), *Introduction to Multiple Time Series Analysis*, Berlin: Springer.

Maheu, J. M. and McCurdy, T. (2002), "Nonlinear Features of Realized FX Volatility," *Review of Economics and Statistics*, 84, 668–681.

Mandelbrot, B., Fisher, A., and Calvet, L. (1997), "A Multifractal Model of Asset Returns," Working Paper, Cowles Foundation, Yale University.

Martens, M., Dijk, D. v., and de Pooter, M. (2004), "Modeling and Forecasting S&P 500 Volatility: Long Memory, Structural Breaks and Nonlinearity," Working Paper, Erasmus University Rotterdam.

Martens, M. and Zein, J. (2004), "Predicting Financial Volatility: High-Frequency Time-Series Forecasts vis-à-vis Implied Volatility," *Journal of Futures Markets*, 24, 1005–1028.

Matheson, J. and Winkler, R. L. (1976), "Scoring Rules for Continuous Probability Distributions," *Management Science*, 22, 1087–1096.

Merton, R. C. (1976), "Option Pricing When Underlying Stock Returns are Discontinuous," *Journal of Financial Economics*, 3, 125–144.

— (1980), "On Estimating the Expected Return on the Market : An Exploratory Investigation," *Journal of Financial Economics*, 8, 323–361.

Mikosch, T. and Stărică, C. (2004), "Nonstationarities in Financial Time Series, the Long-range Dependence, and the IGARCH Effects," *The Review of Economics and Statistics*, 86, 378–390.

Mittnik, S., Paolella, M. S., and Rachev, S. T. (1998), "Unconditional and Conditional Distributional Models for the Nikkei Index," *Asia-Pacific Financial Markets*, 5, 99–128.

— (2000), "Diagnosing and Treating the Fat Tails in Financial Return Data," *Journal of Empirical Finance*, 7, 389–416.

Müller, U. A., Dacorogna, M. M., Davé, R. D., Olsen, R. B., Pictet, O. V., and von Weizsäcker, J. E. (1997), "Volatilities of Different Time Resolutions - Analyzing the Dynamics of Market Components," *Journal of Empirical Finance*, 4, 213–239.

Oomen, R. C. (2004), "Modelling Realized Volatility When Returns are Serially Correlated," Working Paper SP II 2004-11, Social Science Research Center, Berlin.

— (2005), "Properties of Bias-Corrected Realized Variance Under Alternative Sampling Schemes," *Journal of Financial Econometrics*, 3, 555–577.

Pan, J. (2002), "The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study," *Journal of Financial Economics*, 63, 3–50.

Parke, W. (1999), "What is Fractional Integration?" *The Review of Economics and Statistics*, 81, 632–638.

Philipov, A. and Glickman, M. E. (2006), "Multivariate Stochastic Volatility via Wishart Processes," *Journal of Business & Economic Statistics*, 24, 313–328.

Pong, S., Shackleton, M. B., Taylor, S. J., and Xu, X. (2004), "Forecasting Currency Volatility: A Comparison of Implied Volatilities and AR(FI)MA Models," *Journal of Banking & Finance*, 28, 2541–2563.

Poterba, J. M. and Summers, L. H. (1986), "The Persistence of Volatility and Stock Market Fluctuations," *The American Economic Review*, 76, 1142–1151.

Protter, P. (2004), *Stochastic Integration and Differential Equations*, vol. 21 of *Applications of Mathematics*, Springer, 2nd ed.

Robinson, P. (2003), *Time Series With Long Memory*, Oxford University Press.

Rosenblatt, M. (1952), "Remarks on a Multivariate Transformation," *Annals of Mathematical Statistics*, 23, 470–472.

Shin, J. (2005), "Stock Returns and Volatility in Emerging Stock Markets," *Journal of Business & Economic Statistics*, 4, 31–34.

Sowell, F. (1992), "Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models," *Journal of Econometrics*, 53, 165–188.

Tauchen, G. (2005), "Stochastic Volatility in General Equilibrium," Working Paper, Duke University.

Taylor, S. J. (1986), *Modelling Financial Time Series*, Chichester: Wiley.

Taylor, S. J. and Xu, X. (1997), "The Incremental Volatility Information in One Million Foreign Exchange Quotations," *Journal of Empirical Finance*, 4, 317–340.

Thomakos, D. D. and Wang, T. (2003), "Realized Volatility in the Futures Markets," *Journal of Empirical Finance*, 10, 321–353.

Verhoeven, P. and McAleer, M. (2004), "Fat Tails and Asymmetry in Financial Volatility Models," *Mathematics and Computers in Simulation*, 64, 351–361.

Voev, V. (2007), "Dynamic Modelling of Large Dimensional Covariance Matrices," *Empirical Economics*, forthcoming.

Voev, V. and Lunde, A. (2007), "Integrated Covariance Estimation using High-frequency Data in the Presence of Noise," *Journal of Financial Econometrics*, 5, 68–104.

Yu, J. (2005), "On Leverage in a Stochastic Volatility Model," *Journal of Econometrics*, 127, 351–372.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005), "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," *Journal of the American Statistical Association*, 100, 1394–1411.

Zhou, B. (1996), "High-Frequency Data and Volatility in Foreign-Exchange Rates," *Journal of Business & Economic Statistics*, 14, 45–52.