Forecasting S&P 500 Daily Volatility using a Proxy for Downward Price Pressure

Marcel P. Visser *

October 14, 2008

Abstract

This paper decomposes volatility proxies according to upward and downward price movements in high-frequency financial data, and uses this decomposition for forecasting volatility. The paper introduces a simple Garch-type discrete time model that incorporates such high-frequency based statistics into a forecast equation for daily volatility. Analysis of S&P 500 index tick data over the years 1988–2006 shows that taking into account the downward movements improves forecast accuracy significantly. The R^2 statistic for evaluating daily volatility forecasts attains a value of 0.80, both for in-sample and out-of-sample prediction.

JEL classification: C22, C53, G10.

Key Words: volatility proxy, downward absolute power variation, log-Garch, volatility asymmetry, leverage effect, SP500, volatility forecasting, high-frequency data.

*Korteweg-de Vries Institute for Mathematics, University of Amsterdam. Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands. Tel. +31 20 5255861. Email: m.p.visser@uva.nl.

1 Introduction

The way of modelling and forecasting financial volatility has evolved substantially over the past decade. Garch and stochastic volatility models based on daily close-to-close returns are the classical time series models for daily volatility. The availability of large amounts of high-frequency data, recording prices tick-per-tick, has led to new ways of looking at volatility. In particular, one may use high-frequency data to compute for each day a measure called realized volatility. It is current practice to model these realized volatilities directly by fitting AR(FI)MA models. These simple ARMA-type models outperform the Garch-type models (based on daily returns) in out-of-sample forecasting, see for instance Andersen, Bollerslev, Diebold, and Labys (2003), and Koopman, Jungbacker, and Hol (2005).

Besides realized volatility there are other useful measures based on high-frequency data. Barndorff-Nielsen and Shephard (2004) estimate the contribution of jumps to daily price variability using the difference of realized volatility and bipower variation, and Andersen, Bollerslev, and Diebold (2007) use this jump component for forecasting. Brandt and Jones (2006) improve forecast accuracy by using the high-low range. Ghysels, Santa-Clara, and Valkanov (2006) find that the absolute power variation predicts volatility well. Engle and Gallo (2006) show that a 3-dimensional multiplicative error model for the daily absolute return, the intraday high-low range, and the realized volatility is useful for forecasting.

One aspect of intraday prices that has received little attention in the literature is the effect of downward price pressure on future volatility. It is a well-known stylized fact of equity market returns that declining prices go hand in hand with rising volatility. See, in the context of Garch models, Nelson (1991), Engle and Ng (1993), and Glosten, Jagannathan, and Runkle (1993). These papers show that a large negative return today tends to raise tomorrow's volatility more than a large positive return. This phenomenon is known as the leverage effect.

The present paper proposes to capture downward price pressure by using high-frequency price movements. We shall sum the downward absolute five-minute returns, thus obtaining a measure termed *downward absolute power variation*, and use this measure for forecasting daily volatility.

The results in the paper are both theoretical and applied. The main theoretical contribution is the introduction of a simple framework for incorporating statistics that use high-frequency data in a Garch-type forecast equation for daily volatility. In the classical Garch model only daily closing prices are used. We are in the position of observing high-frequency price movements, so it is natural to extend the Garch framework, and replace the daily re-

turn by a proxy based on high-frequency data. Proxies that come to mind are for instance the realized volatility, the bipower variation, and the downward absolute power variation; a precise characterization of the proxies that we allow in the volatility equation is given below. Naturally, if volatility today is high, then these proxies tend to be large, and by the volatility equation this leads to high volatility tomorrow. So, formally, we have a stochastic system where each proxy in the volatility equation contributes to volatility persistence. We shall analyse this system and obtain easy-to-verify stationarity conditions on the parameters of the volatility equation, ensuring stability of the system.

The main empirical contribution of the paper is to show a clear effect of high-frequency downward price pressure as a driving force of S&P 500 index volatility over 1988–2006, and to demonstrate its use in volatility forecasting. In a specification with several explanatory variables the downward absolute power variation has the most pronounced contribution to tomorrow's volatility, whereas the upward absolute power variation adds hardly any explanatory power. We find that measuring downward price pressure by high-frequency data improves forecast accuracy. Specifications that include the downward absolute power variation significantly outperform specifications that do not, both for in-sample and out-of-sample prediction. The Mincer-Zarnowitz R^2 for evaluating daily volatility forecasts yields a value 0.80.

There are alternative ways to capture downward price pressure. Barndorff-Nielsen, Kinnebrock, and Shephard (2008) propose to decompose the realized variance, which is given by the sum of squared intraday returns, into upward and downward components (called semivariances). They discuss the relation of these components to quadratic variation, and in line with our empirical results they find improved log-likelihood values for Garch(1,1) models that include the downward component. We shall also discuss the downward realized variance in our empirical analysis.

The remainder of the paper is organized as follows. Section 2 provides the theoretical framework. Section 3 introduces the downward absolute power variation and evaluates the in-sample fit of the models, and Section 4 presents the out-of-sample forecasting results. Section 5 contains our main conclusions. Appendix A describes the data; Appendix B provides mathematical details on stationarity and invertibility.

2 Accounting for Intraday Price Movements in a Daily Garch Model

2.1 Continuous Time Extensions of Discrete Time Models

Only a decade ago, researchers of financial volatility would typically be analyzing a series of daily close-to-close returns r_n. A commonly applied model for these returns is the Garch(1,1) system, which consists of a return equation and a volatility equation,

$$\begin{aligned} r_n &= \sigma_n Z_n, & (1)\\ \sigma_n^2 &= \kappa + \alpha r_{n-1}^2 + \beta \sigma_{n-1}^2. & (2) \end{aligned}$$

Here, the Z_n are iid, mean zero, unit variance innovations, and κ, α, β are positive parameters. For stationarity one may impose the condition $\alpha + \beta < 1$.

Nowadays we are in the fortunate position of having data on the price movements over the entire trading day. In formal terms, one observes for each trading day n a process $R_n(\cdot)$, the continuous time log-return process for that day. This immediately raises questions of model consistency. Are the intraday return processes $R_n(\cdot)$ consistent with the daily returns in Garch(1,1)? How does one incorporate the processes $R_n(\cdot)$ into this system?

For ease of notation we normalize the trading day to the unit time interval. A basic model for the intraday price movements is the scaled Brownian motion,

$$R_n(u) = \sigma_n W_n(u), \qquad 0 \le u \le 1, \qquad (3)$$

where intraday time u advances from zero to one, see for instance Taylor (1987), and Brandt and Jones (2006). The standard Brownian motion $W_n(\cdot)$ captures intraday price movements, whereas σ_n represents daily volatility and is constant over the day. The present paper adopts the following generalization of equation (3). We allow for an arbitrary process $\Psi_n(\cdot)$, yielding the intraday extension of equations (1–2),

$$\begin{aligned} R_n(u) &= \sigma_n \Psi_n(u), \qquad 0 \le u \le 1, & (4)\\ \sigma_n^2 &= \kappa + \alpha r_{n-1}^2 + \beta \sigma_{n-1}^2, & (5) \end{aligned}$$

where $R_n(1) \equiv r_n$, and $\Psi_n(1) \equiv Z_n$. Specifically, the sample path of $\Psi_n(\cdot)$ is right-continuous and has left limits, one has the standardization $\mathbb{E}\Psi_n^2(1) = 1$, and the sequence of processes $\Psi_n(\cdot)$ is iid. Equation (4) reflects a *scaling model* for the return process over the day.

While this framework does not impose severe constraints on the daily price process, it does allow us to use high-frequency data in modelling daily volatility, as we shall see below.

2.2 Inserting Proxies Into a Log-Garch Volatility Equation

This section presents the basic volatility equation of the paper. Let us first have a closer look at the Garch(1,1) volatilities. The volatility equation (5) states that volatility today (σ_n) is a function of volatility yesterday (σ_{n-1}) and what happened yesterday (reflected by r_{n-1}). In particular, a large price change yesterday, yields a high volatility today (if $\alpha > 0$). In view of the daily return process $R_n(\cdot)$, as it appears in the intraday extension (4–5), it is insightful to rephrase the volatility equation as

$$\sigma_n^2 = \kappa + f(R_{n-1}) + \beta\sigma_{n-1}^2.$$

In the classical situation one uses only the daily returns r_{n-1}, $r_{n-2}, \ldots$, so the statistic $f(R_{n-1})$ is limited to functions of the close-to-close return r_{n-1}; in the case of Garch(1,1)

$$f(R_{n-1}) = R_{n-1}^2(1) \equiv r_{n-1}^2.$$

Given the price movements over the course of the day, $R_{n-1}(\cdot)$, there are many possible statistics that make use of this information. One could use a statistic that gives a good measurement of yesterday's volatility; another possibility is to focus on particular aspects of the sample path of R_{n-1}, such as jumps, or the role of the downward price movements.

A number of statistics based on high-frequency data have appeared in the literature. A commonly applied statistic is the *realized volatility* (RV), see for instance Barndorff-Nielsen and Shephard (2002), and Andersen *et al.* (2003). The statistic RV is frequently used as a proxy for volatility, and is given by the square root of the realized variance. The daily realized variance $RV_n^2(\Delta)$ is the sum of the squared returns over intervals of length Δ, so

$$RV_n(\Delta) = \left(\sum_{k=1}^{1/\Delta} r_{n,k}^2 \right)^{1/2}. \tag{6}$$

For ease of notation, and without loss of generality, we adopt the convention that $1/\Delta$ is an integer. The intraday returns on day n are given by

$$r_{n,k} = R_n(k\,\Delta) - R_n(\,(k-1)\,\Delta). \tag{7}$$

Other statistics are the intraday high-low range (e.g. Parkinson, 1980), and the sum of absolute returns (see Barndorff-Nielsen and Shephard, 2003, 2004). All these statistics have the property of positive homogeneity: if the process $R_n(\cdot)$ is multiplied by a factor $\alpha \geq 0$, then so is the statistic:

$$H(\alpha R_n) = \alpha H(R_n), \qquad \alpha \geq 0. \tag{8}$$

The present paper allows any positive and positively homogeneous statistic. In two recent papers (de Vilder and Visser, 2008, and Visser, 2008), we study this type of statistic and refer to both the random variable H_n,

$$H_n \equiv H(R_n),$$

as well as the functional H as *proxies*.[1] The present paper uses the proxy H_{n-1} as a driver of volatility by incorporating it in the volatility equation. So volatility today depends on volatility yesterday, and a proxy H_{n-1} that reflects specific aspects of yesterday's trading. In particular, the empirical analysis below pays attention to the role of the downward price movements in forecasting volatility. We shall see that including proxies with the scaling property (8) leads to a tractable model for daily volatility.

We incorporate the proxy H_{n-1} into a logarithmic volatility equation; for strictly positive H one may adapt the Garch(1,1) volatility equation as follows:

$$R_n(u) = \sigma_n \Psi_n(u), \qquad 0 \leq u \leq 1, \tag{9}$$

$$\log(\sigma_n) = \kappa + \alpha \log(H_{n-1}) + \beta \log(\sigma_{n-1}), \tag{10}$$

where κ, α and β are real-valued parameters. The system (9–10) constitutes the basic model of this paper; we shall refer to it as the *log-Garch* model. The volatility equation (10) is new, but it has the same interpretation as the classical Garch(1,1) equation for σ_n^2: volatility today (σ_n) depends on volatility yesterday (σ_{n-1}), and on what happened yesterday (reflected by H_{n-1}). If the proxy H_{n-1} is large, then volatility today is large (if $\alpha > 0$).

The use of the log of volatility, $\log(\sigma_n)$, yields easy-to-verify stationarity conditions, as Section 2.3 shows. It also ensures positivity of the volatility process: one does not need to impose on the parameters positivity constraints that may be violated in practice. Early

[1] These papers use H_n as a proxy for σ_n, show that one may improve Garch parameter estimation using proxies, and show how to optimize proxies.

accounts of the use of the logarithm in Garch models are Geweke (1986) and Pantula (1986). Both propose a log-Garch model. Their log-Garch models are similar to equation (10), but apply the logarithm to the squared daily return r_{n-1}^2. That approach is not feasible in practice since daily returns may be zero. The system (9–10) does not suffer from this drawback, as our proxies H_n shall be strictly positive.

2.3 Stationarity for the Log-Garch Model

The log-Garch model admits easy-to-verify stationarity conditions. Since a proxy H_n is linear in σ_n, by $H_n = \sigma_n H(\Psi_n)$, the log of a strictly positive proxy satisfies

$$\log(H_n) = \log(\sigma_n) + U_n, \tag{11}$$

where the $U_n \equiv \log(H(\Psi_n))$ are iid random variables. Inserting relation (11) into the volatility equation (10) one obtains

$$\log(\sigma_n) = \kappa + (\alpha + \beta)\log(\sigma_{n-1}) + \eta_n, \tag{12}$$

where the $\eta_n \equiv \alpha U_{n-1}$ are iid innovations. Equation (12) is simply an autoregressive process of order one (AR(1)) for $\log(\sigma_n)$ with decay parameter $\alpha+\beta$, and mean $(\kappa+\mathbb{E}\eta_n)/(1-\alpha-\beta)$. If η_n has a finite second moment the AR(1) equation is well known to admit a stationary solution if

$$|\alpha + \beta| < 1. \tag{13}$$

More generally one may consider log-Garch(p, q) models that incorporate $j = 1, \ldots, d$ proxies:

$$\log(\sigma_n) = \kappa + \sum_{i=1}^{p}\sum_{j=1}^{d} \alpha_i^{(j)} \log(H_{n-i}^{(j)}) + \sum_{i=1}^{q} \beta_i \log(\sigma_{n-i}), \tag{14}$$

$$= \kappa + \sum_{i=1}^{m} (\bar{\alpha}_i + \beta_i)\log(\sigma_{n-i}) + \eta_n, \tag{15}$$

where $m \equiv \max\{p, q\}$, and $\eta_n \equiv \sum_{i=1}^{p}\sum_{j=1}^{d} \alpha_i^{(j)} U_{n-i}^{(j)}$. Here, $\bar{\alpha}_i \equiv 0$ for $i > p$ and $\beta_i \equiv 0$ for $i > q$. Equation (15) represents an AR(m) process, but is non-standard since the innovations η_n are not independent. The term η_n is similar to an MA(p) component, but is non-standard since it is in general not a moving average of iid innovations if $p > 1$ and $d > 1$. As we show in

the appendix, one may establish stationarity by looking at the AR-polynomial: equation (15) has a unique stationary ergodic solution if the characteristic AR-polynomial $\phi(z)$,

$$\phi(z) = 1 - (\bar{\alpha}_1 + \beta_1)\, z - \ldots - (\bar{\alpha}_m + \beta_m)\, z^m, \tag{16}$$

has only roots outside the unit circle.[2] By the triangle inequality it is sufficient that

$$\sum_{i=1}^{m} |\bar{\alpha}_i + \beta_i| < 1.$$

For details on stationarity, and invertibility, see Appendix B. Invertibility is important, as it ensures that the volatility σ_n can be obtained from observed information.

2.4 Quasi Maximum Likelihood

One may estimate the parameters of a Garch model by the method of maximum likelihood. The traditional approach to Garch parameter estimation is to determine the likelihood by assuming that the daily returns r_n are conditionally Gaussian with mean zero and variance σ_n^2. If the true conditional distribution is not Gaussian, the maximizer of the Gaussian likelihood may still be consistent and asymptotically normal, with adjusted standard errors. It is then called a quasi maximum likelihood estimator (QMLE). For Garch(p, q) processes the QMLE has recently been shown to be consistent and asymptotically normal (Berkes, Horvath, and Kokoszka, 2003); for many other Garch processes consistency and asymptotic normality of the QMLE are open problems. In our empirical analysis we proceed by simply computing the QMLE and providing the Bollerslev and Wooldridge (1992) QML standard errors.

A likelihood for the daily returns r_n does not make use of the information contained in the high-frequency data observed during the course of the day. It is intuitively clear that the use of high-frequency data by means of a suitable volatility proxy of the type given in Section 2.2 may improve the efficiency of parameter estimation: Visser (2008) provides the formal details[3] and introduces a log-Gaussian QMLE; the empirical analysis below uses the log-Gaussian QMLE for parameter estimation. We illustrate the principle for the log-Garch(1,1) model.

[2] This condition excludes non-causal stationary solutions, see Brockwell and Davis (1991).

[3] The details are for the classical Garch(1,1) model, though the principle applies widely to Garch-type models.

First one has to pick a volatility proxy $H^{(0)}$ for which to determine the likelihood function. This does not have to be the same proxy as the proxy H that appears in the volatility equation (10); all that is required is positivity, and positive homogeneity: if σ_n satisfies a log-Garch(1,1) model then the proxy $H_n^{(0)} = \sigma_n\, H^{(0)}(\Psi_n)$ satisfies

$$\begin{aligned}\log(H_n^{(0)}) &= \log(\sigma_n) + U_n^{(0)},\\ &= \kappa_H + \alpha\log(H_{n-1}) + \beta\log(\sigma_{n-1}) + \lambda\varepsilon_n,\end{aligned}$$

where $\kappa_H = \kappa + \mathbb{E}U_n^{(0)}$, λ is the standard deviation of $U_n^{(0)}$, and ε_n is the standardized version of $U_n^{(0)}$, yielding a mean zero, unit variance iid sequence. The conditional mean and variance functions of $\log(H_n^{(0)})$ are

$$\mu_n(\theta) = \kappa_H + \alpha\log(H_{n-1}) + \beta\log(\sigma_{n-1}), \qquad \text{and} \qquad h_n(\theta) = \lambda^2,$$

where $\theta = (\kappa_H, \alpha, \beta, \lambda)$. The QMLE $\hat\theta_N$ is the maximizer of the Gaussian likelihood determined as if

$$\log(H_n^{(0)})|\mathcal{F}_{n-1} \overset{d}{\sim} \mathcal{N}\big(\,\mu_n(\theta),\, h_n(\theta)\,\big),$$

where $\mathcal{F}_{n-1}$ represents observable information up until yesterday. One may use the usual Bollerslev and Wooldridge (1992) QML covariance matrix to obtain empirical standard deviations for the parameter estimates.

3 Full-Sample Analysis

This section provides an in-sample analysis of the daily volatility of the S&P 500 index over the years 1988–2006, a total of 4575 trading days. For a description of the data see Appendix A. Section 3.1 introduces the downward absolute power variation as a measure for downward price pressure. Section 3.2 analyses the explanatory power of the downward absolute power variation in a log-Garch model specification, based on the full sample.

3.1 Downward Price Pressure and Volatility

Before starting to use high-frequency data, let us briefly gain insight in the need for volatility proxies that use intraday price movements to forecast daily volatility. There is a voluminous literature on Garch models based on daily returns alone. One message from this literature for

the empirical modelling of the daily volatility of equity indices and stocks, is the importance of including a *leverage effect*. The leverage effect refers to an asymmetry in the return-volatility relationship: declining prices typically go hand in hand with rising volatility, as already noted by Black (1976) and Christie (1982). More precisely one may distinguish between a leverage effect and a volatility feedback effect; the leverage effect then refers to declining prices that cause volatility, and the volatility feedback effect refers to rising volatility that causes declining prices.[4] The analysis of Bollerslev, Litvinova, and Tauchen (2006) using S&P 500 index five-minute returns strongly suggests that the leverage effect is the more important of the two. A commonly used Garch model that takes into account the leverage effect is the GJR(1,1) model (Glosten, Jagannathan, and Runkle, 1993), which weighs positive and negative returns differently. Estimation of this model on the S&P 500 data yields[5]

$$\begin{array}{lllllllll} \sigma_n^2 = & 1.34\,e\text{-}6 & +0.005 & |r_{n-1}|^2 & +0.096 & |r_{n-1}^-|^2 & +0.936 & \sigma_{n-1}^2, \\ & (3.02\text{e-}7) & (0.009) & & (0.014) & & (0.007) & \end{array} \tag{17}$$

where $r_n^- = \min\{r_n, 0\}$ reflects *downward price pressure*. In accordance with the literature the estimates reflect that a downward price move yesterday tends to intensify volatility today, more so than an upward price move. If only daily returns are available the GJR(1,1) model is hard to beat, see for instance Hansen and Lunde (2005) and Awartani and Corradi (2005), and is quite successful at describing the in-sample returns. This is confirmed by Figures 1(a1) and (a2). Part (a1) depicts the first fifty autocorrelations of the absolute returns, which decay slowly and are significant at all lags. For Garch models, the absolute returns standardized by volatility, $|r_n|/\sigma_n$, are iid. Indeed, Figure 1(a2) shows that the estimated GJR(1,1) volatilities successfully remove the autocorrelation structure, leaving only residual autocorrelations of irregular size and sign.

If high-frequency data are available one may use more efficient proxies to evaluate the volatilities $\hat{\sigma}_n$. In the scaling model of Section 2.1 proxies standardized by volatility,

$$H_n/\sigma_n,$$

[4] An economic explanation for the leverage effect is that a lower stock price increases financial leverage, which entails a larger risk. The volatility feedback effect may be caused by investors demanding a higher expected return, thus lower prices, in the case of increased volatility. Ang, Chen, and Xing (2006) look at downside risk of stocks from a global market perspective, and argue that investors demand a risk premium for bearing downside risk if this risk has a positive correlation with the downside risk of the market portfolio.

[5] Estimation by Gaussian QMLE using the daily returns r_n for $n = 1, \ldots, 4575$, where the first 30 days do not contribute to the likelihood. QML standard errors in parentheses.

form an iid sequence. Figure 1(c2) shows that the GJR(1,1) volatilities do not remove the autocorrelation structure of the five-minute realized volatility $RV5$, where $RV5$ denotes $RV(\Delta = 5 \text{ min.}, 81 \text{ intervals})$. A similar observation applies to the graphs of the high-low range hl and the proxy $H^{(\hat{w})}$ in Figures 1(b2) and (d2), where the proxy $H^{(\hat{w})}$ combines the sum of the ten-minute highs, the sum of the ten-minute lows[6] and the sum of the ten-minute absolute returns as

$$H_n^{(\hat{w})} = (RAV10HIGH_n)^{1.04}(RAV10LOW_n)^{0.72}(RAV10_n)^{-0.76}, \tag{18}$$

which is a good proxy for S&P 500 volatility, see de Vilder and Visser (**?**).

We shall incorporate the intraday price movements in a log-Garch model for σ_n. A natural generalization of the absolute return $|r_n|$ as a proxy for daily volatility is the sum of the absolute returns over successive intervals of length Δ, yielding the *absolute power variation* RAV,

$$RAV_n(\Delta) = \sum_{k=1}^{1/\Delta} |r_{n,k}|,$$

where as before $1/\Delta$ is assumed to be an integer. The intraday returns $r_{n,k}$ on day n are given by (7). The absolute power variation is a good predictor of daily volatility, outperforming the standard realized volatility, see Ghysels, Santa-Clara and Valkanov (2006), and Forsberg and Ghysels (2007). For a discussion of the theoretical properties of RAV for semimartingales, see Barndorff-Nielsen and Shephard (2003, 2004).

Likewise, a sensible proxy for downward price pressure is the sum of the negative returns, $r_{n,k}^- = \min\{r_{n,k}, 0\}$, yielding a novel proxy termed the *downward absolute power variation*

$$RAV_n^-(\Delta) = \sum_{k=1}^{1/\Delta} |r_{n,k}^-|,$$

One may now decompose the absolute power variation as

$$RAV_n(\Delta) = RAV_n^-(\Delta) + RAV_n^+(\Delta),$$

where RAV^+ is the sum of the positive returns. The proxies RAV^- and RAV^+ are positively

[6]The ten-minute high is obtained by the difference of the maximum of $R_n(\cdot)$ and the starting value of $R_n(\cdot)$ over the ten-minute interval in question. The lows are obtained similarly, and made positive by taking absolute values.

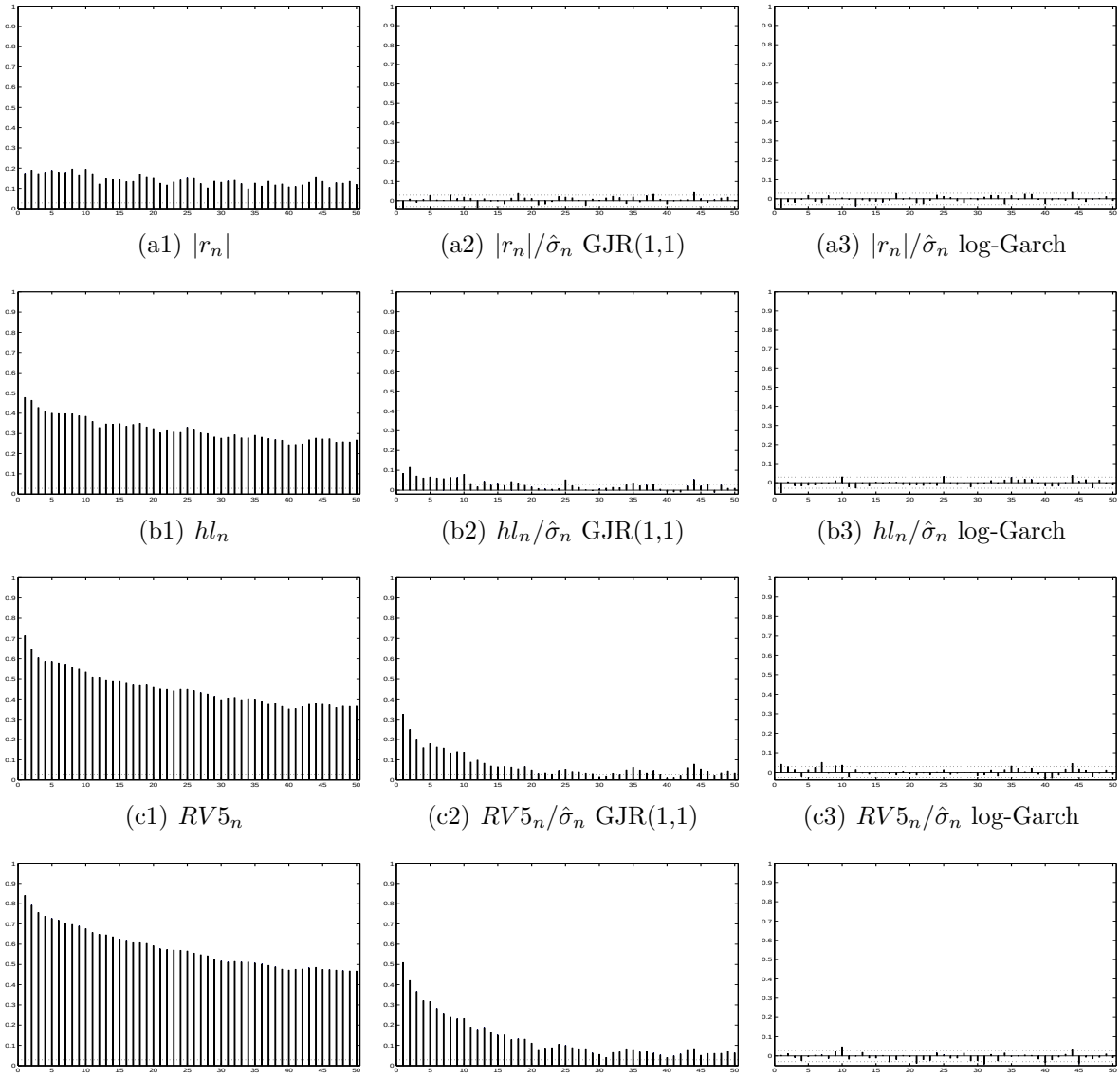

Figure 1: Autocorrelations of four proxies for the days $n = 31, \ldots, 4575$; before and after standardization by σ_n. The volatility proxies change from top to bottom: daily absolute return, intraday high-low range, five-minute realized volatility, and $H^{(\hat{w})}$ as in (18). From left to right different standardization. Leftmost: no standardization. Middle: standardization by GJR(1,1) volatilities $\hat{\sigma}_n$ (equation (17)). Rightmost: standardization by log-Garch where $\hat{\sigma}_n$ uses intraday based volatility proxies (equation (19)). The dotted lines give the standard 95% confidence bounds, $(\pm)2/\sqrt{N}$.

homogeneous; below we shall analyse their use in a log-Garch model.

3.2 A log-Garch Model for the S&P 500 Volatility

In empirical applications volatility processes are typically associated with slowly decaying autocorrelations. One way to deal with the slow decay is to apply a long memory model.[7] We deal with the memory structure by incorporating volatility measurements over the past week and the past month. Such a combination of shorter and longer volatility horizons has been successfully employed in heterogeneous volatility models, such as the HAR-RV specifications for realized volatility in Corsi (2004) and Andersen, Bollerslev, and Diebold (2007), and the HARCH model in Müller *et al.* (1997), which ascribes the relevance of such components to the coexistence of market participants with different trading horizons. In particular we use the weekly and monthly logarithmic moving averages

$$H_{n,\text{Week}}^{(\hat{w}),\log} = \frac{1}{5}\sum_{i=0}^{4}\log(H_{n-i}^{(\hat{w})}), \quad \text{and} \quad H_{n,\text{Month}}^{(\hat{w}),\log} = \frac{1}{22}\sum_{i=0}^{21}\log(H_{n-i}^{(\hat{w})}),$$

where $H_n^{(\hat{w})}$ is given by (18), and 22 is the typical number of trading days in a month.

We specify a log-Garch model that is autoregressive of order one ($q = 1$). The volatility equation includes four kinds of volatility indicators (with parameters $\alpha^{(i)}$, $i = 1, \ldots, 4$):

$$\begin{aligned}\log(\sigma_n) &= \kappa + \alpha^{(1)} H_{n-1,\text{Week}}^{(\hat{w}),\log} + \alpha^{(2)} H_{n-1,\text{Month}}^{(\hat{w}),\log} + \alpha^{(3)}\log(hl_{n-1}) + \alpha^{(4)}\log(RAV5_{n-1}^{-}) \\ &\quad +\beta\log(\sigma_{n-1}),\end{aligned} \tag{19}$$

where $RAV5^{-}$ denotes $RAV^{-}(\Delta = 5$ min., 81 intervals), and hl denotes the intraday high-low range. The top three rows of Table 1 give the full-sample parameter estimates and in parentheses the standard errors and t-values. The estimation uses the log-Gaussian quasi-likelihood for $H_n^{(\hat{w})}$, see Section 2.4. All parameters are highly significant with t-values far outside the 95% region $(-2, 2)$. The estimate $\hat{\beta} = 0.34$ is much smaller than the typical values around 0.9 for traditional Garch; much of the volatility persistence is already captured by the explanatory variables. Volatility over the past week ($\alpha^{(1)}$) and over the past month ($\alpha^{(2)}$) are of similar importance. In line with Engle and Gallo (2006) we find that the high-low range ($\alpha^{(3)}$) has explanatory power in addition to other high-frequency measures of volatility. *The most striking effect is the positive and highly significant effect* $\alpha^{(4)}$ *for the downward absolute power variation* $RAV5_{n-1}^{-}$. The downward price movements appear an important driver of the volatility process.

[7] See for instance the log-ARFIMA model for realized volatility in Andersen *et al.* (2003).

subsample	$\alpha^{(1)}$	$\alpha^{(2)}$	$\alpha^{(3)}$	$\alpha^{(4)}$	β
full	0.166	0.141	0.105	0.214	0.341
	(0.027)	(0.016)	(0.008)	(0.011)	(0.031)
	(6.17)	(8.67)	(12.63)	(19.75)	(11.12)
1st	0.160	0.131	0.120	0.172	0.331
	(0.060)	(0.038)	(0.017)	(0.024)	(0.068)
	(2.66)	(3.46)	(7.10)	(7.04)	(4.84)
2nd	0.126	0.106	0.117	0.164	0.390
	(0.058)	(0.034)	(0.017)	(0.023)	(0.067)
	(2.18)	(3.16)	(6.84)	(7.29)	(5.86)
3rd	0.156	0.133	0.085	0.282	0.284
	(0.051)	(0.033)	(0.017)	(0.021)	(0.055)
	(3.06)	(4.02)	(5.06)	(13.61)	(5.20)
4th	0.142	0.058	0.095	0.221	0.465
	(0.041)	(0.024)	(0.014)	(0.019)	(0.052)
	(3.44)	(2.37)	(6.60)	(11.91)	(8.88)

Table 1: Log-Garch (eq. (19)) parameter estimates based on log-Gaussian QML. The full sample is split into four subsamples. QML standard errors and t-values in parentheses. The estimation in each subsample uses all observations to determine the volatilities $\hat{\sigma}_n$, but leaves the first 30 days out of the likelihood.

Would the parameter $\alpha^{(4)}$ have been as dominant if we had included RAV or RAV^+ in the specification? The inclusion of RAV leads to a small increase in $\hat{\alpha}^{(4)}$ to 0.242 with a t-value 13.56, whereas the parameter value for RAV is slightly negative, -0.056, with a t-value -1.98. The inclusion of RAV^+ yields similar results. This confirms the relevance of distinguishing between upward and downward price movements, and provides further evidence of a pronounced effect of downward price pressure.[8] One could alternatively capture downward price pressure by the downward five-minute realized volatility[9] $RV5^-$, cf. equation (6). Indeed, if we replace $RAV5^-$ by $RV5^-$, we find that the coefficient for $RV5^-$ is large and significant: 0.210 with a t-value of 17.3. It is also interesting to see what happens if we include both $RAV5^-$ and $RV5^-$: we observe a small increase in the parameter $\hat{\alpha}^{(4)}$ for $RAV5^-$ to 0.234 (t-value 7.45), while the parameter value for $RV5^-$ is slightly negative, -0.023, with an insignificant t-value -0.66. Though the quality of the specification that uses $RV5^-$ instead of $RAV5^-$ does not decrease much (the likelihood decreases by roughly 34 points),

[8] Our model (19) does not include RAV, or RAV^+, since their coefficients are small, and change sign in subsamples.

[9] For a discussion of the downward realized volatility, and its relation to quadratic variation, see Barndorff-Nielsen, Kinnebrock, and Shephard (2008). One theoretical difference between summing absolute and squared returns, i.e. $RAV(\Delta)$ or $RV(\Delta)$, is that in the context of semimartingales with a finite activity jump process the measure RV includes the jumps as $\Delta \downarrow 0$, whereas RAV does not (after appropriate scaling). See Barndorff-Nielsen and Shephard (2004).

the parameter estimates favour $RAV5^-$. To check for the effect of $RAV5^-$ over separate time periods, Table 1 gives the parameter estimates for four subsamples spanning the full sample ($n = 1, \ldots, 1143$ and 1144:2287, 2288:3431, 3432:4575). We find that $RV5^-$ does not significantly add explanatory power in any of the subsamples. Moreover, in each subsample the downward absolute power variation is the predominant explanatory variable in (19). We also contrast low with high volatility periods. As a low volatility period we use the days 1003 to 2003 (the four years 1992–1995), and as a high volatility period the days 2600–3700 (the period 1998–05–26 to 2002–11–19). The estimated coefficient $\alpha^{(4)}$ is larger in the high volatility period (as the 2nd and 3rd subsamples in Table 1 also suggest), and $RAV5^-$ is the most pronounced variable in both periods.

The downward absolute power variation and the downward realized volatility are special cases ($r = 1$ and $r = 2$) of the downward r-power variation ($r > 0$),

$$RPV_n^-(\Delta) = \left(\sum_{k=1}^{1/\Delta} |r_{n,k}^-|^r \right)^{1/r}.$$

It is natural to ask for which power r the downward power variation yields the largest likelihood. To this purpose we reestimate the model (19) where RAV^- is replaced by the five-minute RPV^- for various powers r. Figure 3.2 shows full-sample log-likelihood values where we vary the power r over [0.5, 3]. The figure suggests that for $r = 1$, i.e. downward *absolute power* variation, one obtains the best fit.

The message of Figure 3.2 is puzzling. Since the upward and downward absolute power variations are the sums of returns, their difference equals the open-to-close return,

$$RAV_n^+(\Delta) - RAV_n^-(\Delta) \equiv R_n(1) - R_n(0). \tag{20}$$

The difference in explanatory power of RAV^+ and RAV^- may accordingly be attributed to the open-to-close return $R_n(1) - R_n(0)$. The direct use of the open-to-close return, upon replacing $\log(RAV5^-)$ in equation (19) by $R_n(1) - R_n(0)$, yields a significantly lower likelihood.[10] The values of RAV^+ and RAV^- are large and increasing for smaller Δ, so by equation (20) the ratio RAV^+/RAV^- tends to one for small intervals: for small sampling intervals RAV^+ and RAV^- are approximately equal.[11] This implies the existence of an op-

[10] One may observe that such a change of specification does not formally fit into the framework of Section 2.3, so we cannot rely on the stationarity condition provided in equation (16).

[11] This argument may be formalized as follows: the sample paths of a continuous martingale are of unbounded variation, so RAV^+ and RAV^- diverge to infinity, whereas their difference is constant.

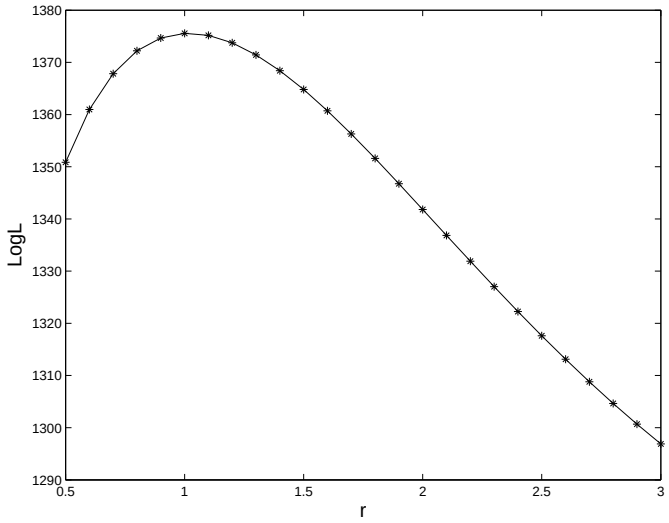

Figure 2: Full-sample log-likelihood values for equation (19) where $RAV5^{-}$ is replaced by the downward r-power variation, $RPV5^{-}$, for various powers r. The power r is varied over [0.5, 3].

timal sampling interval of positive length $\Delta > 0$ on which to base the downward absolute power variation. The log-likelihood based on RAV^{-} for two, five, ten, and thirty-minute intervals is 1372, 1376, 1368, and 1329. It is maximal for the Δ =five-minute sampling interval.

One may wonder whether other variables contribute to equation (19). We test this for a few commonly applied proxies, by separately including the five-minute realized volatility $RV5_{n-1}$, the ten-minute realized range (the square root of the sum of the squared 10-minute high-low ranges, see e.g. Martens and van Dijk, 2007), and the proxy $H_{n-1}^{(\hat{w})}$. None of these measures yields a significant t-value. One may test for the significance of a separate jump component by simultaneously including the five-minute realized volatility and the square root of the five-minute realized bipower variation, where the bipower variation is given by $\sum_{k=2}^{1/\Delta} |r_{n,k}||r_{n,k-1}|$. The bipower variation does not include jumps, at least asymptotically for $\Delta \downarrow 0$, under fairly mild regularity conditions. Whereas Andersen, Bollerslev, and Diebold (2007) find that the bipower variation contributes significantly to the volatility equation, we obtain insignificant t-values. This insignificance is in line with the property that the sum of absolute values (instead of squared values) is robust to jumps (again asymptotically), i.e. jumps in the price process have a relatively minor contribution to the volatility indicators in equation (19).

Table 2 lists the in-sample performance of the six specifications for the volatility σ_n.

The first two columns concern the standard Garch(1,1) and GJR(1,1) models based on daily returns. The third and fourth columns extend the GJR(1,1) model by including the five minute (downward) absolute power variation. The third column, for instance, represents the model

$$\sigma_n^2 = \kappa + \alpha^{(1)} r_{n-1}^2 + \alpha^{(2)} (r_{n-1}^-)^2 + \alpha^{(3)} (RAV5_{n-1}^-)^2 + \beta \sigma_{n-1}^2.$$

The last two columns give the results for the log-Garch model (19) without ($\alpha^{(4)} \equiv 0$) and with the downward absolute power variation $RAV5^-$. The estimation of each specification uses the log-Gaussian quasi-likelihood for $H_n^{(\hat{w})}$. The first row gives the full-sample log-likelihoods.

criterion	Garch(1,1)	GJR(1,1)	GJR(1,1) $+RAV5_{n-1}^-$	GJR(1,1) $+RAV5_{n-1}^-$ $+RAV5_{n-1}$	log-Garch no $RAV5_{n-1}^-$ ($\alpha^{(4)} \equiv 0$ in (19))	log-Garch (eq. (19))
LogL	329.61	419.20	1123.24	1141.06	1145.25	1375.55
$R^2_{\log RV5}$	0.595	0.610	0.679	0.680	0.666	0.689
$R^2_{\log H}$	0.678	0.691	0.773	0.775	0.775	0.797

Table 2: In-sample model comparison for full sample, $n = 31, \ldots, 4575$. Model comparison criteria are the log-likelihood and the R^2's from the logarithmic Mincer-Zarnowitz regression (21), where the volatility proxy is either $RV5$ or $H^{(\hat{w})}$. The two leftmost columns concern the classical Garch(1,1) and GJR(1,1) models. The 3rd and 4th columns extend GJR(1,1) by (downward) absolute power variation. The final two columns concern the log-Garch model (eq. 19) without and with downward absolute power variation. The estimation of each specification uses the log-Gaussian quasi-likelihood for $H_n^{(\hat{w})}$, uses all observations ($n = 1, \ldots, 4575$) to determine the volatilities $\hat{\sigma}_n$, but leaves the first 30 days out of the likelihood.

If the Gaussian density used to determine the likelihood is in fact the true innovations density, then one may use the likelihood ratio statistic (LR) for comparing likelihoods. In the case of QML estimation one may use the QML likelihood ratio statistic given by

$$LR_{qml} = \frac{4}{\mathrm{var}(\varepsilon^2)} (L_1 - L_0),$$

where ε denotes the quasi standard-Gaussian innovation (see Section 2.4), L_1 and L_0 are Gaussian likelihood values, and the factor $4/\mathrm{var}(\varepsilon^2)$ replaces the conventional factor 2 for

the standard LR statistic. The statistic LR_{qml} has the usual chi-square asymptotics,[12] see Busch (2005). Using the full-sample residuals of the log-Garch model (19) we estimate $4/\text{var}(\varepsilon^2) \approx 1.26$. Comparing the first two likelihoods, one sees that the GJR(1,1) model clearly outperforms the standard Garch(1,1) model. This is due to its additional term $(r_{n-1}^-)^2$. The inclusion of the more refined measure $RAV5^-$ greatly raises the log-likelihood value from 419 to 1123.[13] The fourth likelihood value confirms that the absolute power variation $RAV5$ contributes only modestly to the specification, reinforcing the relative importance of the downward price movements captured by $RAV5^-$. The final two columns concern the log-Garch specification, and show that also if one accounts for week and month effects the downward price pressure is a highly powerful predictor of one-day ahead volatility. In addition to likelihood values, Table 2 gives the coefficients of determination R^2 for two Mincer-Zarnowitz (1969) forecast regressions, based on the log of the standard five-minute realized volatility $\log(RV5_n)$ and $\log(H_n^{(\hat{w})})$. The regression equation is given by

$$\log(\text{proxy}_n) = a + b\log(\hat{\sigma}_n) + \varepsilon_n, \tag{21}$$

where we either set $\text{proxy}_n = RV5_n$, or $\text{proxy}_n = H_n^{(\hat{w})}$. In the evaluation of volatility forecasts one has to deal with the complication that the volatility is not observed. So even if one has perfect forecasts, $\hat{\sigma}_n \equiv \sigma_n$, a value $R^2 = 1$ is not achieved, since $RV5_n \neq \sigma_n$ in general. Of course, the larger the R^2 of the regression (21), the larger the predictive ability of $\hat{\sigma}_n$. The full-sample R^2's show the same pattern as the likelihood values: specifications that include the proxy $RAV5^-$ have larger in-sample forecast accuracy. The log-Garch specification outperforms the other specifications, attaining a value of $R^2 \approx 0.80$.

Finally, let us return to Figure 1. From top to bottom, the four rightmost subfigures depict the autocorrelation structure of the proxies $|r|$, hl, $RV5$, and $H^{(\hat{w})}$ after standardization by the log-Garch volatility estimates $\hat{\sigma}_n$. The log-Garch specification (19) outperforms the GJR(1,1) model in removing the autocorrelation structure for each of the four volatility proxies.

[12] Formally, for nested models.

[13] Including $RAV5^-$ in the Garch(1,1) specification yields a log-likelihood 1118, so taking account of the negative daily return r^- adds hardly any explanatory power in the presence of $RAV5^-$.

4 Out-of-Sample Volatility Forecasts

The main practical requirement for a volatility model is that it should be able to forecast volatility (Engle and Patton, 2001). The ultimate test of forecast accuracy is an out-of-sample forecast comparison. We shall apply a number of out-of-sample forecasting criteria to the different models. More specifically, our aim is to gain insight in the out-of-sample predictive ability of $RAV5^{-}$. We shall consider the six specifications of Table 2, and refer to these models as M1 to M6. The forecasts use parameter estimates from rolling samples with a fixed sample size of 1000 days. For each specification we thus generate out-of-sample forecasts

$$\hat{\sigma}_n, \qquad \text{for } n = 1001, \ldots, 4575.$$

The parameter estimates are obtained by log-Gaussian QMLE based on observations $n - 1000, \ldots, n-1$. For each model the likelihood is determined for the proxy $H^{(0)} = H^{(\hat{w})}/c$, where c simply rescales $H^{(\hat{w})}$ such that $\log(H_n^{(0)})$ has the same full-sample average ($n = 1, \ldots, 4575$) as the log five-minute realized volatilities $\log(RV5_n)$. This ensures that we can compare the forecasts $\hat{\sigma}_n$ with $RV5$. To evaluate predictive accuracy we compare the forecasts with two measures of daily volatility: $\log(RV5_n)$, and $\log(H_n^{(0)})$. We use these measures in forecast regressions, $\log(\text{proxy}_n) = a + b\log(\hat{\sigma}_n) + \varepsilon_n$, see (21) above. Unbiasedness corresponds to $a = 0$ and $b = 1$. We shall also compare the regression R^2's with those obtained from in-sample forecasts.

Table 3 gives estimates of the forecast regression coefficients as well as heteroscedasticity and autocorrelation adjusted t-statistics for testing $a = 0$ and $b = 1$. All estimated intercepts $\hat{a}$ are positive, which is (partially) offset by the slopes $\hat{b}$ that all are larger than one (since $\log(RV5_n) < 0$). In the case of the GJR(1,1) model and its extensions with $RAV5$ and $RAV5^{-}$ the t-values lie outside the 95% confidence region $(-2, 2)$, indicating a significant departure from unbiasedness; for the other specifications they are not significant.

The first two rows of Table 4 give the R^2's for the out-of-sample forecast regressions. For comparison, the bottom two rows of Table 4 provide values for predictive accuracy by giving forecast equation R^2's for *in-sample* predictions based on parameter estimates over the period $n = 1001, \ldots, 4575$ (cf. the full-sample values in Table 2). The out-of-sample forecasts are practically as good as the in-sample forecasts, which suggests that the observed high predictive accuracy is not merely an in-sample artifact, or the result of overfitting. Consistent with full-sample analysis the specifications that include a measure for downward

	$\log(RV5)$		$\log(H^{(0)})$	
	intercept	slope	intercept	slope
M1	0.218	1.045	0.188	1.038
	(1.92)	(1.97)	(1.85)	(1.86)
M2	0.219	1.046	0.164	1.033
	(2.13)	(2.18)	(1.73)	(1.73)
M3	0.277	1.057	0.215	1.043
	(4.79)	(4.80)	(3.93)	(3.88)
M4	0.262	1.054	0.203	1.041
	(4.60)	(4.61)	(3.80)	(3.75)
M5	0.056	1.012	0.039	1.008
	(0.87)	(0.91)	(0.80)	(0.76)
M6	0.100	1.021	0.067	1.013
	(1.85)	(1.90)	(1.62)	(1.55)

Table 3: Out-of-sample forecast regression intercepts and slopes for the regression (21), where $\hat{\sigma}_n$, $n = 1001, \ldots, 4575$, are out-of-sample volatility forecasts. The forecasts use parameter estimates from moving windows of 1000 days. Newey-West t-values for $a = 0$ and $b = 1$ in parentheses. The models M1 to M6 are those of Table 2.

		M1	M2	M3	M4	M5	M6
out-of-sample	$R^2_{\log(RV5)}$	0.641	0.664	0.721	0.720	0.704	0.728
	$R^2_{\log(H^{(\hat{w})})}$	0.726	0.744	0.806	0.806	0.802	0.823
in-sample	$R^2_{\log(RV5)}$	0.649	0.672	0.727	0.727	0.708	0.732
	$R^2_{\log(H^{(\hat{w})})}$	0.727	0.747	0.807	0.807	0.803	0.824

Table 4: Forecast regression R^2's for the regression (21), and $\hat{\sigma}_n$, $n = 1001, \ldots, 4575$. The out-of-sample forecasts correspond to those of Table 3. The in-sample forecasts are produced by estimating the parameters over the sample $n = 971, \ldots, 4575$, leaving the first 30 days out of the likelihood. The models M1 to M6 are those of Table 2.

price pressure have larger R^2's than those without downward price pressure.

Finally, Tables 5 and 6 provide for each pair of models a Diebold and Mariano (1995) and West (1996) (DMW) test for the predictive superiority of one model over the other. First, define model i's forecast errors (using a proxy for the actual volatility σ_n)

$$e_{i,n} \equiv \log(\text{proxy}_n) - \log(\hat{\sigma}_{i,n}).$$

Better forecasts have smaller mean squared errors (MSE). One may test for the superiority of model i over model j by testing the significance of the difference in MSE, as given by the

t-test for the $\mu_{i,j}$ coefficient in the regression

$$e_{j,n}^2 - e_{i,n}^2 = \mu_{i,j} + \varepsilon_n, \tag{22}$$

where $\mu_{i,j} > 0$ supports superiority of model i. In both tables the bottom row of each table gives the RMSE, setting the proxy to either $RV5_n$ or $H_n^{(0)}$. The first five rows give the t-statistics, computed using standard-errors with the Newey-West adjustment for heteroscedasticity and autocorrelation. The log-Garch specification (19) outperforms the other models. In all cases the inclusion of $RAV5_n^-$ as a measure for downward price pressure significantly improves forecast accuracy.

log realized volatility: $\log(RV5_n)$

model	M1	M2	M3	M4	M5	M6
M1		-5.28	-9.14	-9.14	-8.12	-10.01
M2			-7.80	-7.79	-5.52	-8.29
M3				0.38	3.41	-2.97
M4					3.43	-3.10
M5						-9.02
RMSE	0.272	0.263	0.241	0.241	0.247	0.237

Table 5: Pairwise tests for superior out-of-sample predictive ability, for $n = 1001, \ldots, 4575$. The (i, j)-th entry in the top five rows gives the t-value for $\mu_{i,j} = 0$ in the regression (22), where proxy_n is the five-minute realized volatility $RV5_n$. A t-value outside the 95% confidence interval $(-2, 2)$ represents statistical significance. The t-value=-5.28 for entry (1,2) supports superiority of model M2 over model M1. The t-values are based on Newey-West adjusted standard errors. The bottom row gives the root mean squared error for each model. The models M1 to M6 are those of Table 2.

log combined proxy: $\log(H_n^{(\hat{w})})$

model	M1	M2	M3	M4	M5	M6
M1		-4.30	-9.49	-9.54	-9.82	-11.28
M2			-8.66	-8.70	-7.92	-10.22
M3				-0.61	0.87	-6.57
M4					0.99	-6.56
M5						-9.10
RMSE	0.222	0.214	0.187	0.187	0.189	0.178

Table 6: Pairwise tests for superior out-of-sample predictive ability as in Table 5, but now with $\text{proxy}_n = H_n^{(0)}$.

5 Conclusions

This paper analyses the effect of downward price pressure, which is measured using high-frequency downward price movements, as a driving force of daily volatility. Its main theoretical contribution is the introduction of a Garch-type discrete time model that incorporates statistics based on high-frequency data into a forecast equation for daily volatility. The paper takes into account intraday price movements in a Garch model for daily volatility. This is achieved by adopting a scaling model for the intraday return process, without imposing severe constraints on the intraday price process. The scaling model offers a continuous time model that yields daily close-to-close returns that satisfy a Garch model. The paper then introduces a Garch-type volatility equation for incorporating statistics such as the realized volatility, and the absolute power variation. The resulting stochastic system leads to easy-to-verify stationarity conditions.

The main empirical result is that the sum of downward absolute five-minute returns (downward absolute power variation), reflecting downward price pressure, is an effective predictor of daily volatility. There is a distinction between explanatory power of upward and downward high-frequency price movements. In a model with several explanatory variables the upward absolute power variation adds hardly any explanatory power, whereas the downward movements are the predominant effect. For the S&P 500 index tick data over 1988–2006, taking into account the downward absolute power variation yields a model that achieves a value $R^2 \approx 0.80$ for evaluating daily volatility forecasts, both for in-sample and out-of sample prediction. The likelihoods for models that use the more general downward r-power variation (summing the r-th power of downward absolute five-minute returns for various r) yield a likelihood plot that is unimodal with a clear maximum at $r = 1$, i.e. the downward *absolute* power variation yields the best fit for the S&P data.

6 Acknowledgment

The author is indebted to Guus Balkema, Chris Klaassen, Remco Peters, and Robin de Vilder for detailed comments and suggestions.

Appendices

A Data

Our data set is the U.S. Standard & Poor's 500 stock index future, traded at the Chicago Mercantile Exchange (CME), for the period 1st of January, 1988 until May 31st, 2006. The data were obtained from Nexa Technologies Inc. (www.tickdata.com). The futures trade from 8:30 A.M. until 15:15 P.M. Central Standard Time. Each record in the set contains a timestamp (with one second precision) and a transaction price. The tick size is \$0.05 for the first part of the data and \$0.10 from 1997–11–01. The data set consists of 4655 trading days. We removed sixty four days for which the closing hour was 12:15 P.M. (early closing hours occur on days before a holiday). Sixteen more days were removed, either because of too late first ticks, too early last ticks, or a suspiciously long intraday no-tick period. These removals leave us with a data set of 4575 days with nearly 14 million price ticks, on average more than 3 thousand price ticks per day, or 7.5 price ticks per minute.

There are four expiration months: March, June, September, and December. We use the most actively-traded contract: we roll to a next expiration as soon as the tick volume for the next expiration is larger than for the current expiration.

B Stationarity and Invertibility

Stationarity and invertibility of a time series are properties that concern the stability of the stochastic system. They play a central role in parameter estimation. Let us first address the question of stationarity. We consider a log-Garch(p, q) model that includes $j = 1, \ldots, d$ proxies, cf. (14–15):

$$\log(\sigma_n) = \kappa + \sum_{i=1}^{p} \sum_{j=1}^{d} \alpha_i^{(j)} \log(H_{n-i}^{(j)}) + \sum_{i=1}^{q} \beta_i \log(\sigma_{n-i}), \tag{23}$$

$$= \kappa + \sum_{i=1}^{m} (\bar{\alpha}_i + \beta_i) \log(\sigma_{n-i}) + \eta_n, \tag{24}$$

where $m \equiv \max\{p, q\}$, $\eta_n \equiv \sum_{i=1}^{p} \sum_{j=1}^{d} \alpha_i^{(j)} U_{n-i}^{(j)}$, and $U_n^{(j)} = \log(H^{(j)}(\Psi_n))$. As before, $\bar{\alpha}_i \equiv 0$ for $i > p$ and $\beta_i \equiv 0$ for $i > q$. Proposition B.1 gives conditions that ensure stationarity. The function $\log^+(\cdot)$ is given by $\log^+(x) = \log(\max\{x, 1\})$. If a random variable

X has a finite r-th moment, $\mathbb{E}|X|^r < \infty$ for some $r > 0$, then $\mathbb{E}\log^+(|X|) < \infty$. Let $(\mathcal{G}_n)$ denote the filtration generated by the processes $\Psi_n(\cdot)$, given by $\mathcal{G}_n = \sigma\{\Psi_n, \Psi_{n-1}, \ldots\}$.

Proposition B.1. *Suppose* $\mathbb{E}\log^+(|\eta_n|) < \infty$*, and define the polynomial* $\phi(z)$*,*

$$\phi(z) = 1 - (\bar{\alpha}_1 + \beta_1)\, z - \ldots - (\bar{\alpha}_m + \beta_m)\, z^m. \tag{25}$$

If all roots of $\phi(z)$ *lie outside the unit circle, then equation (23) admits a unique stationary solution* $(\log(\sigma_n))$*. The stationary solution* $\log(\sigma_n)$ *is ergodic, and is* $\mathcal{G}_{n-1}$*-measurable for all* n*. Moreover, if* $\mathbb{E}|\eta_n|^r < \infty$ *for some* $r > 0$*, then* $\log(\sigma_n)$ *has a finite* r*-th moment.*

Proof. First, the sequence $U_{d,n} \equiv (U_n^{(1)}, \ldots, U_n^{(d)})$, $n \in \mathbb{Z}$, is iid, hence stationary ergodic. The sequence η_n is stationary ergodic, since it is a causal transformation of the stationary ergodic $U_{d,n}$ (Straumann and Mikosch, Proposition 2.5, 2006).

One may write equation (24) in matrix form. Let A^T denote the transpose of A. Let us define an m-dimensional system where $Y_n = (\log(\sigma_n), \ldots, \log(\sigma_{n-m+1}))^T$ and $B_n = \big(\eta_n, 0, \ldots, 0\big)^T$. Equation (24) may now be expressed as

$$Y_n = AY_{n-1} + B_n,$$

where (B_n) is a stationary ergodic sequence, and the $m \times m$ matrix A is given by

$$A = \begin{pmatrix} (\bar{\alpha}_1 + \beta_1) & \ldots & & & (\bar{\alpha}_m + \beta_m) \\ 1 & 0 & \ldots & & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & 1 & 0 \end{pmatrix}.$$

The eigenvalues of A are central to the existence of a stationary solution. As is frequently used in standard ARMA theory, the largest absolute eigenvalue $|\lambda_i|$ (i.e. the spectral radius) of A is smaller than one if the polynomial (25) has only roots outside the unit circle. For a non-stochastic matrix the top-Lyapunov exponent equals the logarithm of the spectral radius, so one may now apply Theorem 1.1 in Bougerol and Picard (1992): Y_n admits the almost sure representation

$$Y_n = \sum_{k=0}^{\infty} A^k B_{n-k},$$

which is the unique stationary solution to (23). Here A^0 represents the identity matrix. The solution Y_n is ergodic, since it is the almost sure limit of a causal transformation of the stationary ergodic sequence (B_n), see Proposition 2.6, Straumann and Mikosch (2006). This proves the claim that $\log(\sigma_n)$ admits a unique stationary and ergodic solution. The B_n, B_{n-1}, ... all are $\mathcal{G}_{n-1}$-measurable. So Y_n is $\mathcal{G}_{n-1}$-measurable, since it is the limit of $\mathcal{G}_{n-1}$-measurable variables.

By ARMA theory one may express $\log(\sigma_n)$ as

$$\log(\sigma_n) = \sum_{i=0}^{\infty} c_i \, \eta_{n-i},$$

where the c_i are given by $\sum_{i=0}^{\infty} c_i z^i = 1/\phi(z)$, and $\sum_{i=0}^{\infty} |c_i| < \infty$, see for instance Brockwell and Davis (1991). By the triangle inequality and dominated convergence one has

$$\mathbb{E}|\log(\sigma_n)|^r \leq \mathbb{E}\left(\lim_{k\to\infty} \sum_{i=0}^{k} |c_i \eta_{n-i}|\right)^r = \lim_{k\to\infty} \mathbb{E}\left(\sum_{i=0}^{k} |c_i \eta_{n-i}|\right)^r.$$

By assumption $\mu_r \equiv \mathbb{E}|\eta_n|^r < \infty$. Applying Minkowski's inequality and the absolute summability of the c_i,

$$\mathbb{E}|\log(\sigma_n)|^r \leq \lim_{k\to\infty} \left(\sum_{i=0}^{k} (\mathbb{E}|c_i \eta_{n-i}|^r)^{1/r}\right)^r = \mu_r \left(\lim_{k\to\infty} \sum_{i=0}^{k} |c_i|\right)^r < \infty.$$

□

Let us now turn to the question of invertibility. Algorithms for estimating parameters and forecasting are typically only effective under invertibility. Consider the log-Garch(1,1) specification (10), and suppose that $\log(\sigma_n)$ is a stationary solution. One does not observe σ_0, and in practice one typically replaces this value by a starting value $\hat{\sigma}_0 > 0$, and simply iterates the recursion

$$\log(\hat{\sigma}_n) = \kappa + \alpha \log(H_{n-1}) + \beta \log(\hat{\sigma}_{n-1}),$$

for $n = 1, \ldots, N$. Following Straumann and Mikosch (2006), we say that the process $\log(\sigma_n)$

is *invertible*[14] if

$$|\log(\hat{\sigma}_n) - \log(\sigma_n)| \xrightarrow{\mathbb{P}} 0, \qquad n \to \infty,$$

i.e. the approximation becomes arbitrarily precise (given the true parameter values). Application of the invertibility definition to the general specification (23) reveals that invertibility concerns only the parameters β. Let the filtration $(\mathcal{F}_n)$ represent the observed information given by the intraday return processes $R_n(\cdot)$, so $\mathcal{F}_n = \sigma\{R_n, R_{n-1}, \ldots\}$.

Proposition B.2. *Let the process* $(\log(\sigma_n))$ *be a stationary solution to the log-Garch equation (23). Define the polynomial* $\phi_\beta(z)$*,*

$$\phi_\beta(z) = 1 - \beta_1 z - \ldots - \beta_q z^q. \tag{26}$$

If $q = 0$ *then* $(\log(\sigma_n))$ *is invertible. If* $q > 0$ *and all roots of* $\phi_\beta(z)$ *lie outside the unit circle then* $(\log(\sigma_n))$ *is invertible. An invertible solution* $\log(\sigma_n)$ *is* $\mathcal{F}_{n-1}$*-measurable for all* n*.*

Proof. If there are no autoregression parameters ($q = 0$), then the approximation scheme is exact, hence the process is invertible, and is $\mathcal{F}_{n-1}$-measurable.
Consider the case $q > 0$. In analogy to the proof of Proposition B.1 define the q-dimensional vector $Y_n = (\log(\sigma_n), \ldots, \log(\sigma_{n-q+1}))^T$, and the d-dimensional vector $B_n = \Big(\log(H_n^{(1)}), \ldots, \log(H_n^{(d)})^T$. By definition, the stationary solution Y_n satisfies the recursion

$$Y_n = AY_{n-1} + \sum_{i=1}^{p} A_i B_{n-i},$$

for all n. Here, the $q \times q$ matrix A and the $q \times d$ matrices A_i are given by

$$A = \begin{pmatrix} \beta_1 & \ldots & & & \beta_q \\ 1 & 0 & \ldots & & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & 1 & 0 \end{pmatrix}, \qquad \text{and} \qquad A_i = \begin{pmatrix} \alpha_i^{(1)} & \ldots & & \alpha_i^{(d)} \\ 0 & 0 & \ldots & 0 \\ \vdots & \ddots & & \vdots \\ & & & \\ 0 & \ldots & & 0 \end{pmatrix},$$

To obtain the reconstruction vector $\hat{Y}_n$, start with arbitrary values n days back, $\log(\hat{\sigma}_0) =$

[14]The usual ARMA invertibility ensures that the ARMA innovations may be expressed in terms of the present and past of the observables; the concept of invertibility here may be seen as a generalization, see Straumann and Mikosch (2006, Section 3.2).

$\hat{y}_0, \ldots, \log(\hat{\sigma}_{-q+1}) = \hat{y}_{-q+1}$, and iterate

$$\hat{Y}_n = A\hat{Y}_{n-1} + \sum_{j=1}^{d} A_j B_{n-j}, \qquad n \geq 1.$$

One has

$$Y_n - \hat{Y}_n = A(Y_{n-1} - \hat{Y}_{n-1}) = \ldots = A^n(Y_0 - \hat{Y}_0), \qquad n = 1, 2, \ldots$$

So $||Y_n - \hat{Y}_n|| \leq ||A^n||_{op}$ almost surely, where $||B||_{op}$ denotes the operator norm for a matrix B, given by $||B||_{op} = \sup_{x \neq 0} \frac{||Bx||}{||x||}$. Let $\rho(A)$ denote the spectral radius of A. One has, in general, $\lim_{n\to\infty} ||A^n||_{op}^{1/n} = \rho(A)$, so

$$\lim_{n\to\infty} ||Y_n - \hat{Y}_n|| \stackrel{a.s.}{=} 0,$$

if $\rho(A) < 1$. In analogy to the proof of Proposition B.1, one has $\rho(A) < 1$ if the roots of $\phi_\beta(z)$ lie outside the unit circle.

In general one could start the reconstruction iteration k days back, and obtain the k-th backward iterate $\hat{Y}_{n,k}$ (in particular, $\hat{Y}_{n,n} \equiv \hat{Y}_n$). Note that $\hat{Y}_{n,k}$ is $\mathcal{F}_{n-1}$-measurable for all $k > 0$. By stationarity $Y_n - \hat{Y}_{n,k} \stackrel{d}{=} Y_k - \hat{Y}_{k,k} = Y_k - \hat{Y}_k$. So invertibility of the Y_k (i.e. $Y_k - \hat{Y}_k$ converges to zero in probability) is equivalent to $Y_n - \hat{Y}_{n,k} \to 0$ in probability for $k \to \infty$. Convergence in probability implies the existence of a subsequence k_i such that

$$Y_n = \lim_{i\to\infty} \hat{Y}_{n,k_i},$$

almost surely, hence an invertible Y_n is $\mathcal{F}_{n-1}$-measurable. □

Remark B.1. *If the conditions of Proposition B.1 hold for $r = 2$ then $\log(\sigma_n)$ is covariance stationary.*

Remark B.2. *If $q = 1$ in Proposition B.2 one has invertibility if $-1 < \beta < 1$.*

Remark B.3. *One may think of parameter configurations that satisfy the conditions for stationarity, but not those for invertibility. An example is a log-Garch(1,1) model with $|\alpha + \beta| < 1$ and $\beta > 1$.*

Remark B.4. *Under the conditions of both Proposition B.1 and Proposition B.2 one has $\mathcal{F}_n \equiv \mathcal{G}_n$. This may be seen by the following arguments. If the conditions of Proposition B.1*

are satisfied, then $\mathcal{F}_n \subset \mathcal{G}_n$, since $R_n(\cdot) = \sigma_n \Psi_n(\cdot)$, and σ_n is $\mathcal{G}_n$-measurable. If the conditions of Proposition B.2 are satisfied, then $\mathcal{G}_n \subset \mathcal{F}_n$, since $\Psi_n(\cdot) = R_n(\cdot)/\sigma_n$, and σ_n is $\mathcal{F}_n$-measurable.

Remark B.5. *It is possible to include positive proxies that are positively homogeneous of a degree $r > 0$ in the volatility equation. The focus on $r \equiv 1$ in the present paper is without loss of generality. Suppose that $\tilde{H}$ is positively homogeneous of degree r. Then $H \equiv (\tilde{H})^{1/r}$ is positively homogeneous of degree 1. Then, $\log(\tilde{H}(R_n)) = r \log(H(R_n))$, so the effect of $\tilde{H}$ may simply be captured by H.*

Remark B.6. *It is fairly easy to extend the log-Garch model by the following class of intraday statistics. Let R_n denote, as before, the daily return process. Consider a statistic $D_n \equiv D(R_n)$ that is positively homogeneous of degree zero,*

$$D(\alpha R_n) = D(R_n), \qquad \alpha \geq 0.$$

Examples of such a statistic are the ratio of two proxies, the ratio of the daily return and the realized volatility, or the time of the intraday high. The statistic D_n satisfies

$$D_n = D(\sigma_n \Psi_n) = D(\Psi_n),$$

so the D_n form an iid sequence. Inclusion of a term $\delta\, D_{n-1}$ in equation (23) only alters the innovation η_n in (24). The conditions for stationarity and invertibility of the log-Garch model, as given by Propositions B.1 and B.2, remain unchanged.

References

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. *The Review of Economics and Statistics*, **89,** number 4, 701–720.

Andersen, T.G., Bollerslev, T., F.X., Diebold and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, **71,** number 2, 579–625.

Ang, A, Chen, J. and Xing, Y. (2006). Downside Risk. *The Review of Financial Studies*, **19,** number 4, 1191–1239.

Awartani, B.M.A. and Corradi, V. (2005). Predicting the volatility of the S&P-500 stock index via GARCH models: the role of asymmetries. *International Journal of Forecasting*, **21,** number 1, 167–183.

Barndorff-Nielsen, O.E., Kinnebrock, S. and Shephard, N. (2008). Measuring downside risk – realised semivariance. CREATES Research Paper 2008-42, University of Aarhus.

Barndorff-Nielsen, O.E. and Shephard, N. (2002). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, **17,** number 5, 457–477.

Barndorff-Nielsen, O.E. and Shephard, N. (2003). Realized power variation and stochastic volatility models. *Bernoulli*, **9,** number 2, 243–65 and 1109–1111.

Barndorff-Nielsen, O.E. and Shephard, N. (2004). Power and Bipower Variation with Stochastic Volatility and Jumps. *Journal of Financial Econometrics*, **2,** number 1, 1–37.

Berkes, I., Horvath, L. and Kokoszka, P. (2003). Garch processes: structure and estimation. *Bernoulli*, **9,** number 2, 201–227.

Black, B. (1976). Studies of stock price volatility changes. In *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics*, pp. 177–181.

Bollerslev, T., Litvinova, J. and Tauchen, G. (2006). Leverage and Volatility Feedback Effects in High-Frequency Data. *Journal of Financial Econometrics*, **4,** number 3, 353–384.

Bollerslev, T. and Wooldridge, J.M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, **11,** number 2, 143–172.

Bougerol, P. and Picard, N. (1992). Strict Stationarity of Generalized Autoregressive Processes. *The Annals of Probability*, **20,** number 4, 1714–1730.

Brandt, M.W. and Jones, C.S. (2006). Volatility Forecasting With Range-Based EGARCH Models. *Journal of Business & Economic Statistics*, **24,** number 4, 470–486.

Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods*. Springer Series in Statistics, second edn. New York: Springer-Verlag.

Busch, T. (2005). A robust LR test for the GARCH model. *Economics Letters*, **88,** 358–364.

Christie, A.C. (1982). The Stochastic Behavior of Common Stock Variances–Value, Leverage and Interest Rate Effects. *Journal of Financial Economics*, **10,** number 4, 407–432.

Corsi, F. (2004). A Simple Long Memory Model of Realized Volatility. SSRN Paper.

de Vilder, R.G. and Visser, M.P. (2008). Ranking and Combining Volatility Proxies for Garch and Stochastic Volatility Models. MPRA paper no. 11001.

Diebold, F.X. and Mariano, R.S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, **13,** number 3, 253–263.

Engle, R.F. and Gallo, G.M. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, **131,** number 1-2, 2–27.

Engle, R.F. and Ng, V.K. (1993). Measuring and Testing the Impact of News on Volatility. *The Journal of Finance*, **48,** number 5, 1749–1778.

Engle, R.F. and Patton, A.J. (2001). What good is a volatility model? *Quantitative Finance*, **1,** number 2, 237–245.

Forsberg, L. and Ghysels, E. (2007). Why Do Absolute Returns Predict Volatility So Well? *Journal of Financial Econometrics*, **5,** number 1, 31–67.

Geweke, J. (1986). Modelling the Persistence of Conditional Variances – Comment. *Econometric Reviews*, **5,** number 1, 57–61.

Ghysels, E., Santa-Clara, P. and Valkanov, R. (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, **131,** number 1-2, 59–95.

Glosten, L.R., Jagannathan, R. and Runkle, D.E. (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance*, **48,** number 5, 1779–1801.

Hansen, P.R. and Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, **20,** number 7, 873–889.

Koopman, S.J., B., Jungbacker 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,** number 3, 445–475.

Martens, M. and van Dijk, D. (2007). Measuring volatility with the realized range. *Journal of Econometrics*, **138,** number 1, 181–207.

Mincer, J. and Zarnowitz, V. (1969). The Evaluation of Economic Forecasts. In *Economic Forecasts and Expectation* (ed. J. Mincer), National Bureau of Economic Research, pp. 3–46. Columbia University Press.

Müller, U.A., Dacorogna, M.M., Davé, R.D., Olsen, R.B., Pictet, O.V. and Weizsäcker, J.E. (1997). Volatilities of Different Time Resolutions – Analyzing the Dynamics of Different Market Components. *Journal of Empirical Finance*, **4,** number 2-3, 213–239.

Nelson, D.B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, **59,** number 2, 347–370.

Pantula, S.G. (1986). Modelling the Persistence of Conditional Variances – Comment. *Econometric Reviews*, **5,** number 1, 71–74.

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

Straumann, D. and T., Mikosch (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. *The Annals of Statistics*, **34,** number 5, 2449–2495.

Taylor, S.J. (1987). Forecasting the volatility of currency exchange rates. *International Journal of Forecasting*, **3,** number 1, 159–170.

Visser, M.P. (2008). Garch parameter estimation using high-frequency data. MPRA paper no. 9076.

West, K.D. (1996). Asymptotic inference about predictive ability. *Econometrica*, **64,** number 5, 1067–1084.