



ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

Journal of Empirical Finance 12 (2005) 445–475

Journal of  
EMPIRICAL  
FINANCE

[www.elsevier.com/locate/econbase](http://www.elsevier.com/locate/econbase)

# Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements<sup>☆</sup>

Siem Jan Koopman<sup>a,b,\*</sup>, Borus Jungbacker<sup>a</sup>, Eugenie Hol<sup>c</sup>

<sup>a</sup>*Department of Econometrics, Free University Amsterdam, The Netherlands*

<sup>b</sup>*Tinbergen Institute Amsterdam, The Netherlands*

<sup>c</sup>*ING Group Credit Risk Management, The Netherlands*

Accepted 28 April 2004

Available online 3 February 2005

## Abstract

The increasing availability of financial market data at intraday frequencies has not only led to the development of improved volatility measurements but has also inspired research into their potential value as an information source for volatility forecasting. In this paper, we explore the forecasting value of historical volatility (extracted from daily return series), of implied volatility (extracted from option pricing data) and of realised volatility (computed as the sum of squared high frequency returns within a day). First, we consider unobserved components (UC-RV) and long memory models for realised volatility which is regarded as an accurate estimator of volatility. The predictive abilities of realised volatility models are compared with those of stochastic volatility (SV) models and generalised autoregressive conditional heteroskedasticity (GARCH) models for daily return series. These historical volatility models are extended to include realised and implied volatility measures as explanatory variables for volatility. The main focus is on forecasting the daily variability of the Standard & Poor's 100 (S&P 100) stock index series for which trading data (tick by tick) of almost 7 years is analysed. The forecast assessment is based on the hypothesis of whether a forecast model is outperformed by alternative models. In particular, we will use superior predictive ability tests to

<sup>☆</sup> An earlier version of this paper was circulated with title “Stock Index Volatility Forecasting with High Frequency Data”. All errors in the paper are ours.

\* Corresponding author. Department of econometrics, Free University Amsterdam, De Boelelaan 1105, NL-1081 HV Amsterdam, The Netherlands.

E-mail address: [s.j.koopman@feweb.vu.nl](mailto:s.j.koopman@feweb.vu.nl) (S.J. Koopman).

URL: <http://www.staff.feweb.vu.nl/koopman/>.

investigate the relative forecast performances of some models. Since volatilities are not observed, realised volatility is taken as a proxy for actual volatility and is used for computing the forecast error. A stationary bootstrap procedure is required for computing the test statistic and its  $p$ -value. The empirical results show convincingly that realised volatility models produce far more accurate volatility forecasts compared to models based on daily returns. Long memory models seem to provide the most accurate forecasts.

© 2004 Elsevier B.V. All rights reserved.

*JEL classification:* C22; C53; G15

*Keywords:* Generalised autoregressive conditional heteroskedasticity model; Long memory model; Realised volatility; Stochastic volatility model; Superior predictive ability; Unobserved components

---

## 1. Introduction

Modelling and forecasting volatility in financial markets has gained much interest in the financial and economic literature. The seminal paper of Engle (1982) has started the development of a large number of so-called historical volatility models in which a time-varying volatility process is extracted from financial returns data. Most volatility models can be regarded as variants of the generalised autoregressive conditional heteroskedasticity (GARCH) models of Bollerslev (1986); see Bollerslev et al. (1994) for a review. A rival class of volatility models is associated with the stochastic volatility (SV) model; see Taylor (1986) and Harvey et al. (1994). The overviews presented in Shephard (1996) and Ghysels et al. (1996) provide an excellent introduction to historical volatility models. A more recent review of volatility models together with an assessment of their forecasting performances is given by Poon and Granger (2003).

Both GARCH and SV models are regularly used for the analysis of daily, weekly and monthly returns. From a theoretical perspective, these models can also be applied to returns data measured at higher frequencies (intraday). However, it is learned from empirical studies that these models cannot accommodate all information in high frequency returns. The initial work of Andersen and Bollerslev (1998) and Barndorff-Nielsen and Shephard (2001) show that realised volatility (a daily volatility measure) as computed by the cumulative sum of squared intraday returns is less subject to measurement error and therefore less noisy. It follows that measurement noise contained in daily squared returns is too high for observing the underlying volatility process. The measurement noise is reduced as the sampling frequency of the return series is increased, although market microstructure effects prevent us from using sampling intervals that are too small; see Andersen et al. (2001b) and Barndorff-Nielsen and Shephard (2001, 2002). These results also justify the earlier work of French et al. (1987), among others. Andersen and Bollerslev (1998) show that daily forecasts of exchange rates based on GARCH models, when evaluated against realised volatility, are far more accurate than had been previously assumed. These findings were subsequently confirmed with regard to stock index data by Blair et al. (2001) and Hansen and Lunde (2005), who examined the predictive accuracy of volatility forecasts based on GARCH models.

Volatility can be extracted from returns data, but it can also be derived from option pricing data in combination with an option pricing model. Early empirical studies by

Latanè and Rendleman (1976), Chiras and Manaster (1978) and Beckers (1981) have indicated that implied volatility, when compared with historical standard deviations, can be regarded as a good predictor of future volatility. Implied volatility is often referred to as the market's volatility forecast and is said to be forward looking as opposed to historical-based methods, which are by definition backward looking. More recent studies by Christensen and Prabhala (1998), Fleming (1998), Blair et al. (2001) and Giot (2003) show that accurate volatility forecasts for returns on stock indices are often based on implied volatility. Moreover, their research strongly suggests that daily returns contain little or no incremental information about future volatility.

In this paper, we investigate the potential gains of different measures of volatility and different ways of modelling these data for the purpose of volatility forecasting. For example, Blair et al. (2001) and Martens (2001) suggest to incorporate realised volatility as an explanatory variable in the variance equation of a daily GARCH model. They found a considerable improvement in the forecasting performance in this way. Another possible explanatory variable for volatility is implied volatility. We will explore this option further by incorporating such explanatory variables in both GARCH and SV models.

Realised volatility can also be modelled directly which is reminiscent of the methods adopted for monthly volatility in a number of earlier studies such as those by French et al. (1987) and Poon and Taylor (1992). The forecasting performance of realised volatility models has been studied, among others, by Andersen et al. (2001a), and Barndorff-Nielsen and Shephard (2004a). In the first paper, it is stressed that long memory features are present in the logarithms of realised volatility, and that the autoregressive fractionally integrated moving average (ARFIMA-RV) model is effective in empirical modelling. The second paper builds on Barndorff-Nielsen and Shephard (2002), where volatility is represented as a continuous time series process, the sum of independent Lévy-driven Ornstein–Uhlenbeck (OU) processes. This approach forms the basis of an unobserved components (UC-RV) model for realised volatility that consists of independent ARMA components with restricted parameters.

The empirical investigation is for the Standard & Poor's 100 (S&P 100) stock index series over the period 6 January 1997 to 14 November 2003 with 1725 trading days. Opening and closure prices for all trading days in the sample are available in this period together with all price quotes within the days (tick by tick). Further, we have obtained the S&P 100-implied volatility index from the Chicago Board Options Exchange Market volatility index (VIX) which is known to be a highly liquid options market. The forecasting performance of various volatility models for the last 525 days of the data set is the focus of the empirical study. We compare the forecasts of ARFIMA-RV, UC-RV, SV and GARCH volatility models; the latter two models are considered with and without explanatory variables. The forecasts are generated by a rolling-window of 1200 observations through the last 525 daily observations. Forecast comparison is based on four different loss functions including the mean squared error (MSE) and the mean absolute error (MAE) statistics. The fact that a particular loss criterion is smallest for a particular model does not provide any information about its forecast superiority in other samples of the data set and in future samples of the data. The results in White (2000) and the important refinements in Hansen (2001) constitute a framework that constructs a formal test for superior prediction ability (SPA) of a benchmark or base model relative to a

set of rival models. Since volatility can never be observed, realised volatility is taken as a proxy for actual volatility and used for determining the forecast error. This may introduce inconsistencies in the ranking of forecast models, but it is argued that the occurrence of such inconsistencies is unlikely in our study. The method of computing the SPA test statistic and its  $p$ -value requires bootstrap samples obtained by, for example, the stationary bootstrap procedure of Politis and Romano (1994). The construction of the test and some details of implementation are discussed.

The findings of this extensive empirical study are presented by reporting a selection of the most interesting results. The maximum likelihood estimates for the coefficients of the considered models are reported for the full sample. Although these estimates are not used for forecasting since all models are reestimated for each rolling window sample (starting from 17 October 2001), the reported estimation results provide insights about the S&P 100 data set and the effectiveness of models to capture volatility information from the data. A selection of the forecasting results is also presented, but most attention is paid to the SPA results. It has become clear that the realised volatility models are overwhelmingly superior, and therefore, making comparisons between, say, GARCH and ARFIMA-RV is not useful. We therefore concentrate on the comparison of models within the two classes of realised volatility models and historical volatility models. It will be concluded that both the ARFIMA-RV and the SV models with realised volatility as the explanatory variable are superior within their classes for the forecasting of S&P 100 volatility.

The remainder of this paper is organised as follows. In the next section, we discuss the available data sets and discuss how returns are constructed and how volatility measures are derived from these returns. In Section 3, we provide some details about the realised volatility models, the UC-RV and ARFIMA-RV models. In Section 4, the historical volatility models are shortly described. The forecastig methodology is discussed in Section 5, and in Section 6, the estimation and forecasting results are presented. Section 7 concludes.

## 2. Stock return data and volatility

### 2.1. Daily and intraday returns

The data for our empirical study consists of Standard & Poor's 100 (S&P 100) stock index price quotes during the period from 6 January 1997 to 14 November 2003, which constitutes a total of  $N=1725$  trading days. The daily return  $R_n$  is defined as the first difference between the 4:00 p.m. closing prices on consecutive trading days, expressed in percentage terms; that is

$$R_n = 100(\ln P_n - \ln P_{n-1}), \quad n = 1, \dots, N, \quad (1)$$

where  $P_n$  is the closing asset price at trading day  $n$ .

In the S&P 100 data set,<sup>1</sup> the price quotes are available for all trading days. We can therefore extract intraday returns for any interval that is usually measured in minutes. Such

<sup>1</sup> The S&P 100 index quotes are obtained from Disk Trading <http://www.is99.com/disktrading/>. Database consists of tick by tick quotes measured in US dollars and concerns the current S&P 100 index (no futures prices).

high-frequency returns are subject to market microstructures (bid-ask bounces, discrete price observations, irregular trading, etc.) which distort the return properties of interest. The 5-min frequency is regarded as the highest frequency at which the effects of market microstructure are not too distorting. Andersen and Bollerslev (1998) also construct their intraday returns using the 5-min frequency. A more complete discussion on the choice of frequency is given by Andersen et al. (2001b). To construct the 5-min return, the last price quote is recorded before the relevant time mark, and the difference is taken between successive log prices; that is

$$R_{n,d} = 100(\ln P_{n,d} - \ln P_{n,d-1}), \quad n = 1, \dots, N, \quad d = 1, \dots, D, \quad (2)$$

where  $P_{n,d}$  is the asset price at trading day  $n$ , at the 5-min mark  $d$ . Note that  $P_{n,0}$  is the opening price of day  $n$ .

The New York Stock Exchange opens at 9:30 a.m. and closes at 4 p.m. EST. A full trading day therefore consists of  $D=78$  intraday returns and one overnight return. Not all trading days will have 78 5-min quotes because the New York Stock Exchange closes early on certain days (for example, Christmas Eve). Another important reason for a different number of 5-min returns is the lapses in trading and in data reporting on some trading days. Various methods are used to deal with such problems. For example, Andersen et al. (2001a) consider linear interpolation through bid-ask quotes in order to obtain a full set of 5-min interval returns. A flexible spline interpolation method is used by Hansen and Lunde (2005).

Following this recent literature, we have used a spline method for the construction of 5-min returns as follows. All price quotes within one trading day are ordered over the trade sequence. The duration between price quotes is measured in minutes and is strictly nonnegative. In the majority of cases multiple price quotes occurs in 1 min. The durations of price quotes after the first quote in a minute are set to zero. The resulting time series of price quotes is effectively a nonequally spaced sequence. A nonparametric spline function is estimated through these high-frequency intraday prices. The spline function through a nonequally spaced time series can be represented as a time-varying state space model; see Durbin and Koopman (2001, Section 3.11). To attain an optimal level of smoothing, the “smoothing parameter” of the spline function is estimated via the maximisation of the (quasi) Gaussian log-likelihood function. The spline function itself is computed by the Kalman filter and an associating smoothing algorithm<sup>2</sup>. The 5-min nodes of the spline are taken as the “observed” prices  $P_{n,d}$  for  $d=0, \dots, D$ . The price  $P_{n,0}$  is defined as the observed opening price which is the first price observed after 9:29 a.m. (even if it takes place at, say, 11:30 a.m.). The closing price  $P_{n,D}$  is then the last price observed before 4:00 p.m. The overnight returns are defined as

$$R_{n,0} = 100(\ln P_{n,0} - \ln P_{n-1,D}), \quad n = 1, \dots, N. \quad (3)$$

In this way, we have constructed 79 intradaily returns for the 1725 trading days in our data set.

<sup>2</sup> These computations are carried out using Ox3.3 (see Doornik (2001)) with use of the SsfPack 3.0 functions SsfGetSpline, SsfLikEx and SsfMomentEstEx (see Koopman et al. (1999)).

## 2.2. Realised volatility

It is generally acknowledged that squared daily returns provide a poor approximation of actual daily volatility. It was first pointed out by [Andersen and Bollerslev \(1998\)](#) that more accurate estimates could be obtained by summing squared intraday returns. They defined realised volatility in the foreign exchange market as the sum of 288 5-min squared returns. If we were to apply this method directly to the stock market, realised volatility would be defined as the sum of the squared overnight and the cumulative squared 5-min intraday returns, so

$$\hat{\sigma}_n^2 = R_{n,0}^2 + \sum_{d=1}^D R_{n,d}^2, \quad n = 1, \dots, N, \quad (4)$$

with  $R_{n,d}$  and  $R_{n,0}$  as defined in Eqs. (2) and (3), respectively, and with  $D=78$ . However, this ignores the fact that the overnight return is a special case. Stock markets, unlike foreign exchange markets, are not open 24 h a day, and the changes in the stock index price during the hours that the stock market is closed are relatively large compared to the 5-min returns observed during trading hours. In order to account for the fact that overnight returns are presumably more volatile than intraday 5-min returns and that a large value for  $R_{n,0}$  will have a pronounced and distorting effect on the realised volatility estimate  $\hat{\sigma}_n^2$ , an alternative realised volatility measures need to be considered by excluding the “noisy” overnight return  $R_{n,0}$ . This will measure the volatility during trading hours as opposed to daily volatility<sup>3</sup>. It is suggested by [Martens \(2002\)](#) to scale the sum of intraday returns by

$$\hat{\sigma}_n^2 = \frac{\hat{\sigma}_{oc}^2 + \hat{\sigma}_{co}^2}{\hat{\sigma}_{oc}^2} \sum_{d=1}^D R_{n,d}^2, \quad (5)$$

with the “open-to-close” sample variance  $\hat{\sigma}_{oc}^2$  and the “close-to-open” sample variance  $\hat{\sigma}_{co}^2$  computed as

$$\begin{aligned} \hat{\sigma}_{oc}^2 &= \frac{10,000}{N} \sum_{n=1}^N (\log P_{n,D} - \log P_{n,0})^2, \\ \hat{\sigma}_{co}^2 &= \frac{10,000}{N} \sum_{n=1}^N (\log P_{n,0} - \log P_{n-1,D})^2. \end{aligned}$$

An alternative sample estimate of the “open-to-close” variance is the average of the realised volatilities of Eq. (4), but without  $R_{n,0}^2$ , from 1 to  $N$  as considered by [Hansen and Lunde \(2002\)](#). However, the current proposed estimates  $\hat{\sigma}_{oc}^2$  and  $\hat{\sigma}_{co}^2$  are consistent with each other since they are both based on sums of  $N$  squared returns. An alternative scaling is considered by [Areal and Taylor \(2002\)](#) who assign different weights to the intraday squared returns with weights depending on variance proportions. [Hansen and Lunde \(2005\)](#) provide a more detailed discussion on the choice of scaling.

<sup>3</sup> [Andersen and Bollerslev \(1997\)](#) and [Andersen et al. \(2001a\)](#) use this definition of realised volatility in their stock market studies.

Other recent methods for the construction of realised volatility include the Fourier method of Malliavin and Mancino (2002), and applied by Barucci and Reno (2002) and Hansen and Lunde (2003), and the filtering method of Corsi et al. (2001).

### 2.3. Implied volatility

The implied volatility index is obtained from the Chicago Board Options Exchange Market Volatility Index<sup>4</sup> (VIX) which is based on a highly liquid options market for the period 6 January 1997 to 14 November 2003. The VIX index is calculated from the midpoint of bid-ask option prices using a binomial method, as described in Harvey and Whaley (1992), which takes account of the level and timing of dividend payments on the underlying S&P 100 stock index. The Black–Scholes model assumption of constant volatility, however, introduces bias into the implied volatility measure. Hull and White (1987) found that the magnitude of the bias in the Black–Scholes model was smallest for near-the-money and close-to-maturity options; also see Feinstein (1995). Therefore, the index is based on implied volatilities of eight nearby-expiry call and put options.<sup>5</sup> The VIX index is then constructed as a weighted average of these implied volatilities in such a way that it represents the implied volatility of a hypothetical at-the-money OEX option with 22 trading days to expiration. Although the VIX represents a biased implied volatility measure, it is still a useful proxy of future volatility for the practical application of volatility forecasting. It mitigates many of the measurement errors which typically contribute to biased implied volatility measures. The VIX is constructed on a daily basis and is denoted by  $s_n^2$  for  $n=1, \dots, N$ .

### 2.4. Descriptive statistics

In Fig. 1, we present graphs of the time series that are used in this empirical study together with their histograms and correlograms of the time series. The summary statistics of the series are given in Table 1.

For the daily return series  $R_n$ , we observe several volatile periods which occurred towards the end of 1997, during the third quarter of 1998, at the beginning and end of 2000 and during the summer of 2002. The three largest shocks to the return process took place in one of these periods and were negative. This largely contributed to the reported negative skewness coefficient for the return series and the large positive skewness coefficient of the squared returns. It is also seen that  $R_n$  and  $R_n^2$  exhibit excess kurtosis.

The realised volatility series  $\hat{\sigma}_n^2$  is presented in Fig. 1 in levels (row 3) and logs (row 4). The distribution of the logged realised volatility appears approximately Gaussian but with a fat tail to the left due to one “inlier” in the summer of 1998. The descriptive statistics in Table 1 for realised volatility mirror earlier findings for stock

<sup>4</sup> The VIX data were extracted from the CBOE on-line database.

<sup>5</sup> The shortest time to maturity is at least 8 days in order to eliminate the increase in implied volatility observed during the option’s last week of trading. The calendar-day implied volatilities of each of the eight options are adjusted to a trading base as in Fleming et al. (1995). The wildcard option implicit in the options is ignored as in Fleming and Whaley (1994).



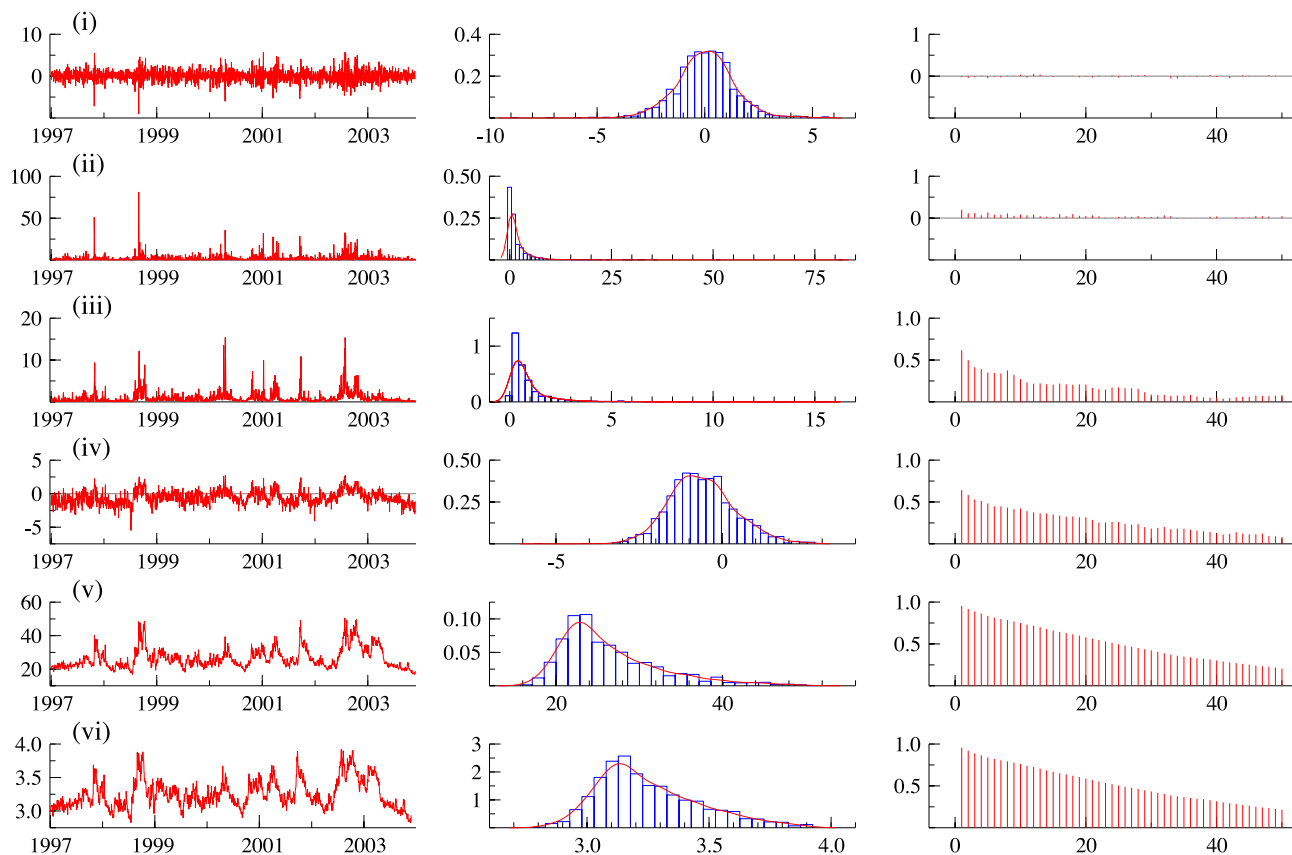


Fig. 1. Six daily time series with observations (first column), histogram (second column) and correlogram (third column): (i) daily returns  $R_n$ ; (ii) squared daily returns; (iii) realised volatility  $\sigma_n^2$ ; (iv) realised volatility in logs; (v) implied volatility  $s_n^2$ ; (vi) implied volatility in logs. The time series are related to the Standard & Poor's 100 stock index for the period from 6 January 1997 to 14 November 2003 (1725 days).



Table 1

Summary statistics of return and volatility time series for the period from 6 January 1997 to 14 November 2003 (number of days is 1725)

	Daily return		Realised volatility		Implied volatility	
	$R_n$	$R_n^2$	$\tilde{\sigma}_n^2$	$\log \tilde{\sigma}_n^2$	$s_n^2$	$\log s_n^2$
Mean	0.020	1.889	0.920	−0.612	26.46	3.253
Standard deviation	1.374	4.058	1.359	0.981	5.998	0.208
Skewness	−0.122	7.918	5.109	0.245	1.266	0.744
Excess kurtosis	5.621	110.8	39.80	0.524	1.482	0.135
Minimum	−8.994	0	0.004	−5.484	16.84	2.834
Maximum	5.702	80.89	15.38	2.733	50.48	3.922

market data, as reported by Andersen et al. (2001a), Areal and Taylor (2002) and Giot and Laurent (2004).

The daily implied volatility (VIX) series  $\tilde{\sigma}_n^2$  in levels (row 5) and logs (row 6) are also included in Fig. 1, and their summary statistics are reported in Table 1. The effects of some large outliers in the sample are illustrated by high values for skewness and excess kurtosis. Various changes in implied volatility can be observed, and they correspond more or less with changes in realised volatilities. However, high volatility periods appear more pronounced in the implied volatility series than in the realised volatility series.

The time series properties are revealed by the sample autocorrelation functions or correlograms which are presented in the third column of the graphs in Fig. 1. Positive serial correlation is present in the squared daily return series but is not strong, as the volatility dynamics are subject to measurement error. This is reduced considerably for realised volatility, and more correlation structure in the series appears as a result, both for levels and logs. Most notably, long memory features are apparent in logged realised volatility. The most persistent correlation structure appears in the implied volatility series.

### 3. Realised volatility models

The spot price of an asset is denoted by  $P(t)$ , and its return is defined as

$$R(t) = \log P(t) - \log P(0), \quad t > 0. \quad (6)$$

Motivated by financial economic theory, the dynamic stochastic process of returns can be represented by the continuous time process

$$dR(t) = \mu(t)dt + \sigma(t)dW(t), \quad t > 0, \quad (7)$$

where  $\mu(t)$  is a drift process,  $\sigma(t)$  is the spot volatility and  $W(t)$  is a standard Brownian motion process. The drift term  $\mu(t)$  is included for completeness but will be of no importance below. For the introduction of various volatility models below, we further assume that the mean and variance of spot volatility are given by

$$E(\sigma^2(t)) = \xi, \quad \text{var}(\sigma^2(t)) = \omega^2.$$

The actual volatility for the  $n$ -th day interval of length  $h$  is then defined as

$$\sigma_n^2 = \sigma^*(hn) - \sigma^*((n-1)h), \text{ where } \sigma^*(t) = \int_0^t \sigma^2(s)ds.$$

### 3.1. Unobserved components model for realised volatility

It has been argued that realised volatility  $\tilde{\sigma}_n^2$  is an accurate estimator of actual volatility  $\sigma_n^2$ ; see, for example, Andersen and Bollerslev (1998). Barndorff-Nielsen and Shephard (2002) provide an excellent discussion of the statistical properties of this estimator and its error  $\sigma_n^2 - \tilde{\sigma}_n^2$ . They also show with a Monte Carlo study that using a model for the spot volatility  $\sigma^2(t)$  can significantly improve estimates of actual volatility. A candidate model for  $\sigma^2(t)$  is based on the superposition of Ornstein–Uhlenbeck (OU) processes  $\tau^j(t)$ ; that is

$$\sigma^2(t) = \sum_{j=1}^J \tau^j(t), \quad d\tau^j(t) = -\lambda_j \tau^j(t)dt + dz_j(\lambda_j t), \quad (8)$$

where  $z_j(t)$  is an independent Lévy process with nonnegative increments (also known as a subordinator) and  $\lambda_j$  is an unknown fixed parameter for  $j=1, \dots, J$  with  $J$  as the number of factors. In this case, the stochastic differential equation defining  $\tau^j(t)$  permits its autocorrelation function to be expressed as

$$\text{corr}(\tau^j(t), \tau^j(t+s)) = e^{-\lambda_j |s|}.$$

By assuming  $E(\tau^j(t)) = w_j \xi$  and  $\text{var}(\tau^j(t)) = w_j \omega^2$ , where  $w_j > 0$  and  $\sum w_j = 1$ , a convenient and intuitively appealing form for the autocorrelation function for  $\sigma^2(t)$  can be derived and is given by

$$\text{corr}(\sigma^2(t), \sigma^2(t+s)) = \sum_{j=1}^J w_j e^{-\lambda_j |s|}.$$

where we refer to Barndorff-Nielsen and Shephard (2001, 2002) for more details on this approach to modelling volatility. Using the expression of the autocorrelation  $\text{corr}(\tau^j(t), \tau^j(t+s))$  and after some calculus, it can be shown that the autocorrelation function of the  $j$ -th component of actual volatility  $\tau_n^j \equiv \int_{(n-1)h}^{nh} \tau^j(t)dt$  is given by

$$\text{corr}(\tau_n^j, \tau_{n+m}^j) = \frac{(1 - e^{-\lambda_j h})^2}{2(e^{-\lambda_j h} - 1 + \lambda_j h)} e^{-\lambda_j h(m-1)}, \quad m = 1, 2, \dots,$$

where  $h$  is the length of the day interval. This convenient result implies that  $\tau_n^j$  permits the ARMA(1,1) representation

$$\tau_{n+1}^j = w_j \xi + \phi_j (\tau_n^j - w_j \xi) + \theta_j \eta_{n-1}^j + \eta_n^j, \quad \eta_n^j \sim \text{WN}(0, \sigma_{\eta}^2), \quad (9)$$

where  $WN(0, \sigma^2)$  refers to a white noise process with zero mean and variance  $\sigma^2$ . It follows that the autoregressive parameter  $\phi_j$  equals  $e^{-\lambda_j h}$ , while Meddahi (2003) shows that

$$\theta_j = \frac{1 - \sqrt{1 - 4\vartheta_j^2}}{2\vartheta_j}, \text{ with } \vartheta_j = \frac{\text{corr}(\tau_n^j, \tau_{n+1}^j) - \phi_j}{(1 + \phi_j^2) - 2\phi_j \text{corr}(\tau_n^j, \tau_{n+1}^j)}, \quad (10)$$

where  $\theta_j$  has the typical value of 0.26; see also Barndorff-Nielsen and Shephard (2004b). The key to modelling realised volatility is the set of results in Barndorff-Nielsen and Shephard (2001, Section 2.3) in which they define the estimation error as  $u_n = \sigma_n^2 - \tilde{\sigma}_n^2$  and where, assuming that the return process (7) is valid, they establish it to be a (weak) white noise sequence with mean zero and variance

$$\sigma_{u,n}^2 = 2D_n \left( (\xi h/D_n)^2 + \sum_{j=1}^J \frac{2w_j \omega^2}{\lambda_j^2} \left( e^{-\lambda_j h/D_n} - 1 + \lambda_j h/D_n \right) \right), \quad (11)$$

with  $D_n$  given by the number of intradaily intervals used to calculate  $\tilde{\sigma}_n^2$ . By assuming that the model for spot volatility is correctly specified, the model for realised volatility is then finally obtained as

$$\tilde{\sigma}_n^2 = \sum_{j=1}^J \tau_n^j + u_n, \quad \tau_n^j \sim \text{ARMA}(1, 1), \quad u_n \sim \text{WN}(0, \sigma_{u,n}^2), \quad (12)$$

where  $\tau_n^j$  and  $\sigma_{u,n}^2$  are given by Eqs. (9) and (11), respectively. The resulting model is in effect an unobserved ARMA components (UC) model.

The unobserved ARMA components model can be represented in the state space formulation, and methods based on the Kalman filter can be used for the quasi-maximum likelihood estimation of fixed parameters such as  $\lambda_j$  and for the signal extraction of spot volatility conditional on realised volatility  $\tilde{\sigma}_n^2$  which takes account of the implied dynamic structure of spot volatility. More importantly, the model framework allows model-based forecasting of tomorrow's spot volatility based on the unobserved ARMA components model. It should be noted that the analysis is based on quasi-maximum likelihood estimation because distributions for  $u_n$  and  $\eta_n^j$  are not established.

### 3.2. Long memory ARFIMA model

In empirical work on realised volatility, it is pointed out that the realised volatility series  $\tilde{\sigma}_n^2$  can be modelled by a Gaussian dynamic process after it is transformed into logs. The dynamic properties of log realised volatility exhibit features known as long memory; that is, the correlogram decays less than exponentially as the lag length increases. The correlograms of the realised volatility series in levels and logs are presented in Fig. 1 to illustrate that this empirical fact also applies to our S&P 100 data set. To model the long memory properties, the autoregressive fractionally integrated moving average (ARFIMA) model is adopted following Andersen et al. (2001a) and Andersen et al. (2003).

The ARFIMA(1, $d$ ,1) model with mean  $\mu$  for realised volatility can be given by

$$(1 - \phi L)(1 - L)^d (\tilde{\sigma}_n^2 - \mu) = (1 + \theta L)\varepsilon_n, \quad (13)$$

where  $L$  is the lag operator ( $L\tilde{\sigma}_n^2 = \tilde{\sigma}_{n-1}^2$ ), coefficients  $d$ ,  $\phi$  and  $\theta$  are fixed and unknown and  $\varepsilon_n$  is Gaussian white noise with mean zero and variance  $\sigma^2$ . The following restrictions on the parameters apply,

$$0 < d < 0.5, \quad |\phi| < 1, \quad |\theta| < 1, \quad \sigma^2 > 0.$$

In the context of volatility modelling, the ARFIMA model for the logs of realised volatility is empirically investigated; for example, by Andersen et al. (2003).

The parameters of the ARFIMA model, including mean  $\mu$ , can be estimated by the method of maximum likelihood; for details, see, for example, Sowell (1992) and Doornik and Ooms (2003). However, the estimation is not without hurdles. It is, for example, pointed out by Brodsky and Hurvich (1999) and Bos et al. (2002) that standard ARMA(1,1) models can also capture long memory features and that, depending on the sample spectrum of the data, not all parameters of an ARFIMA(1, $d$ ,1) can be empirically identified from the data. This typically applies to the case of realised volatility. For example, Andersen et al. (2003) estimate the  $d$  parameter using a log-periodogram estimator and fix  $d$  as estimated in their long memory vector autoregressive model of order 1. However, we estimate the  $d$  parameter simultaneously with other coefficients of the model. Since different ARFIMA model specifications produce rather similar results, we consider the ARFIMA(1, $d$ ,0) model in the empirical study<sup>6</sup>.

Forecasting can be carried out by extrapolating the series in which the correlation structure implied by the estimated ARFIMA model is taken into account. Details of how these computations can be implemented elegantly for ARFIMA models are given by Doornik and Ooms (2003).

#### 4. Daily time-varying volatility models

In this section, we discuss daily time-varying volatility models where volatility is explicitly modelled as the second moment of daily returns. This class of so-called historical volatility models is well established and includes stochastic volatility (SV) and generalised autoregressive conditional heteroskedasticity (GARCH) classes of models. In addition to standard formulations, we also consider extensions to both models with realised and implied volatility measures incorporated in the variance equation.

##### 4.1. Stochastic volatility model

The stochastic volatility model is deduced from the continuous time process for the returns (6). By discretising the continuous process (7) at daily intervals and by assuming

<sup>6</sup> The required computations are implemented using the ARFIMA package of Doornik and Ooms (2001) within the programming environment of Ox; see Doornik (2001).

an autoregressive process for log-volatility, we obtain the stochastic volatility model for daily returns as given by

$$\begin{aligned} R_n &= \mu + \sigma_n \varepsilon_n, \quad \varepsilon_n \sim \text{NID}(0, 1), \quad n = 1, \dots, N, \\ \sigma_n^2 &= \sigma^{*2} \exp(h_n), \\ h_{n+1} &= \phi h_n + \sigma_\eta \eta_n, \quad \eta_n \sim \text{NID}(0, 1), \quad h_1 \sim \text{NID}\left(0, \sigma_\eta^2 / \{1 - \phi^2\}\right), \end{aligned} \quad (14)$$

where  $R_n$  is computed as in Eq. (1), and constant  $\mu$  is treated as fixed and zero in our empirical study. The choice of a first-order autoregressive process for the daily log-volatility  $h_n$  is common. Note that some weak serial correlation in  $R_n^2$  exists as confirmed by the first correlogram of Fig. 1. The actual volatility  $\sigma_n^2$  is modelled by the product of a scaling factor  $\sigma^{*2} > 0$  and the exponential of the stochastic process  $h_n$ . The persistence parameter  $\phi$  of the autoregressive process is restricted to be positive and smaller than one to ensure stationarity. We further assume that  $\varepsilon_n$  and  $\eta_n$  are mutually uncorrelated, contemporaneously and at all lags, and with  $h_1$ . More theoretical details about the SV model are discussed by Shephard (1996).

The nonlinear relation between the unobserved log-volatility  $h_n$  and the observed dependent variable  $R_n$  does not allow the computation of the likelihood by linear methods such as the Kalman filter. However, the likelihood function of the SV model can be constructed using simulation methods such as the ones developed by Shephard and Pitt (1997) and Durbin and Koopman (1997). These developments allow the exact maximum likelihood estimation of parameters of the SV model using Monte Carlo importance sampling techniques; see also Sandmann and Koopman (1998).

For the SV model, we can express the likelihood function as

$$L(\psi) = p(y|\psi) = \int p(y, \theta|\psi) d\theta = \int p(y|\theta, \psi) p(\theta|\psi) d\theta, \quad (15)$$

where  $y = (y_1, \dots, y_N)'$  and

$$\psi = (\phi, \sigma_\eta, \sigma^*)', \quad \theta = (h_1, \dots, h_N)'.$$

An efficient way of evaluating such expressions is by using importance sampling; see Ripley (1987, Chap. 5). A simulation device is required to sample from an importance density  $\tilde{p}(\theta|y, \psi)$  which we prefer to be as close as possible to the true density  $p(\theta|y, \psi)$ . It is common to take a conditional Gaussian density  $g(\theta|y, \psi)$  as the importance density since it is relatively straightforward to sample from it using simulation smoothers such as the ones developed by de Jong and Shephard (1995) and Durbin and Koopman (2002). Guidelines for the construction of Gaussian importance sampling devices for the SV model are discussed in Lee and Koopman (2004). It can be shown that the Monte Carlo estimate of the likelihood function can be computed as

$$\hat{L}(\psi) = L_G(\psi) D^{-1} \sum_{j=1}^D d_j, \quad d_j = \frac{p(y|\theta^j, \psi)}{g(y|\theta^j, \psi)},$$

where  $\theta^j$  is a draw (realisation) of  $g(\theta|y,\psi)$  as obtained from the simulation smoother for  $j=1, \dots, D$ . The likelihood function can be computed for a given value of parameter vector  $\psi$  and can be numerically maximised with respect to  $\psi$  by using the same source of random deviates for computing  $\hat{L}(\psi)$  for the different values of  $\psi$ . In this way, maximum likelihood estimates of  $\psi$  are obtained.<sup>7</sup>

The one-step ahead daily volatility forecast for the SV model is computed as

$$\hat{\sigma}_{N+1}^2 = \hat{\sigma}^{*2} \exp(\hat{h}_{N+1|N} + 0.5p_{N+1|N}), \quad (16)$$

where  $\hat{\sigma}^{*2}$  is the maximum likelihood estimate of  $\sigma^{*2}$ ,  $\hat{h}_{N+1|N}$  is the estimator of  $h_{N+1}$  given all  $N$  observed daily returns and  $p_{N+1|N}$  is its mean squared error. The latter two values are computed by

$$\hat{h}_{N+1|N} = \frac{\sum_j d_j h_{N+1|N}(h^j)}{\sum_j d_j}, \quad p_{N+1|N} = \frac{\sum_j d_j p_{N+1|N}(h^j)}{\sum_j d_j},$$

where  $h_{N+1|N}(\theta^j)$  and  $p_{N+1|N}(\theta^j)$  are obtained from the Kalman filter applied to the approximating model that is used as part of the Gaussian importance sampling device and is based on the draw  $\theta^j$  from  $g(\theta|y,\psi)$ . Further details of this approach are discussed in [Durbin and Koopman \(2000\)](#).

The SV model can be extended by the inclusion of an explanatory variable in the log-variance equation for  $h_n$  in Eq. (14); that is

$$h_n = \phi h_{n-1} + \gamma(1 - \phi)Lx_{n-1} + \sigma_\eta \eta_n, \quad n = 1, \dots, N, \quad (17)$$

where  $x_n = \ln \tilde{\sigma}_n^2$  for realised volatility or  $x_n = \ln s_n^2$  for implied volatility. An equivalent specification for model (17) is given by

$$h_n = \gamma x_{n-1} + \eta_n^*, \quad \eta_n^* = \phi \eta_{n-1}^* + \sigma_\eta \eta_n,$$

which is a regression model with autoregressive disturbances. Although  $\phi$  is restricted to be positive for model (14), for model (17), the standard stationary condition  $-1 < \phi < 1$  applies. For other volatility models, similar extensions have been considered; see [Martens \(2001\)](#) and [Blair et al. \(2001\)](#), among others. Replacing the Eq. (17) for  $h_n$  in Eq. (14) does not affect the nonlinear relationship between daily return  $R_n$  and unobserved volatility  $h_n$ . Therefore, the estimation and forecasting methods for the SV model, as described earlier, can be applied straightforwardly. The regression coefficient  $\gamma$  in Eq. (17) is estimated by numerically maximising the simulated likelihood function with respect to  $\gamma$  and jointly with respect to the other coefficients. One-step ahead forecasts for the extended model are obtained using the same methods as for the SV model, although it requires the value of  $x_N$ .

<sup>7</sup> The SV model is estimated using a program written in the Ox language of [Doornik \(2001\)](#) using SsfPack by [Koopman et al. \(1999\)](#). The Ox programs can be obtained from <http://staff.feweb.vu.nl/koopman/sv/>.

#### 4.2. Generalised autoregressive conditional heteroskedasticity model

An alternative to the SV model and more common in its use of modelling time-varying volatility is the generalised autoregressive conditional heteroskedasticity (GARCH) model which in its most simplest form is given by

$$R_n = \sigma_n \varepsilon_n, \quad \varepsilon_n \sim \text{NID}(0, 1), \quad n = 1, \dots, N,$$

$$\sigma_n^2 = \omega + \alpha R_{n-1}^2 + \beta \sigma_{n-1}^2, \quad (18)$$

with parameter restrictions  $\omega > 0$ ,  $\alpha \geq 0$ ,  $\beta \geq 0$  and  $\alpha + \beta < 1$ . The parameter vector  $\psi$  consists of  $\omega$ ,  $\alpha$  and  $\beta$  in this case. The persistence of daily volatility is measured by  $\alpha + \beta$  in GARCH models. More lags of  $R_n^2$  and  $\sigma_n^2$  can be added to the volatility equation of the GARCH model, but in empirical studies, the model (18) has proven to work effectively in forecasting volatility; see, for example, Andersen and Bollerslev (1998) and Hansen and Lunde (2005). The time-varying variance is formulated conditional on the lagged observed return  $R_{n-1}$ , and therefore, the variance  $\sigma_n^2$  is known conditional on past observed returns and parameter vector  $\psi$ . The predictive observation density is given by

$$R_{n+1} | R_1, \dots, R_n, \psi \sim \text{NID}(0, \omega + \alpha R_n^2 + \beta \sigma_n^2).$$

The prediction error decomposition allows the construction of the likelihood function in a straightforward way; see Bollerslev (1986). When the assumption of normality for  $\varepsilon_n$  is dropped, estimation can be referred to as quasi-maximum likelihood. Estimation procedures for the GARCH model are implemented in many standard econometric software packages<sup>8</sup>. Interesting surveys on GARCH models are written by Bollerslev et al. (1992), Bera and Higgins (1993) and Bollerslev et al. (1994). Finally, it follows directly that the one-day ahead volatility forecast at time  $N$  can be computed as

$$\hat{\sigma}_{N+1}^2 = \hat{\omega} + \hat{\alpha} R_N^2 + \hat{\beta} \hat{\sigma}_N^2, \quad (19)$$

where  $\hat{\omega}$ ,  $\hat{\alpha}$  and  $\hat{\beta}$  denote maximum likelihood estimates of  $\omega$ ,  $\alpha$  and  $\beta$ , respectively.

The daily GARCH model can also be extended to include an explanatory variable by incorporating it in the variance equation, so the volatility process  $\sigma_n^2$  in Eq. (18) can be rewritten as

$$\sigma_n^2 = \omega + \alpha R_{n-1}^2 + \beta \sigma_{n-1}^2 + \gamma s_{n-1}^2, \quad (20)$$

where  $s_n^2$  represents implied volatility but can be exchanged with realised volatility  $\tilde{\sigma}_n^2$ . We note that the inclusion of  $s_n^2$  in the volatility Eq. (20) is different from its inclusion in Eq. (17), and therefore, the regression coefficients have different implications for forecasting. Most econometric packages have options to specify explanatory variables for the volatility Eq. (20) because estimation of these coefficients is relatively straightforward. As for the

<sup>8</sup> In this paper, a program for estimating the parameters of the GARCH model is written in Ox, the programming environment of Doornik (2001), using functions of the G@RCH package of Laurent and Peters (2002).



standard GARCH model, all information for the calculation of the one-day ahead volatility forecast  $\hat{\sigma}_{N+1}^2$  is available at time  $N$  and is given by

$$\hat{\sigma}_{N+1}^2 = \hat{\omega} + \hat{\alpha}R_N^2 + \hat{\beta}\hat{\sigma}_N^2 + \hat{\gamma}s_N^2, \quad (21)$$

where  $\hat{\gamma}$  is the maximum likelihood estimate of  $\gamma$ .

## 5. Forecasting methodology

In the next section, the results are presented of an out-of-sample forecasting study in which we compare the forecasting performances of the various volatility models described in Sections 3 and 4. This forecasting study is carried out as follows. The volatility models are estimated 525 times based on 525 samples of 1200 observations. The first sample starts 6 January 1997 and ends 16 October 2001. A forecast of volatility is generated for 17 October 2001 based on the estimated model for the first sample. The second sample, starting 7 January 1997 and ending 17 October 2001, is used to forecast the volatility of 18 October 2001 based on the estimated model for the second sample. These estimation and forecasting steps can be repeated 525 times for the available sample from 6 January 1997 to 14 November 2003. More specifically, we produce the 525 one-day ahead forecasts  $\hat{\sigma}_m^2$  where

$$\hat{\sigma}_m^2 \text{ is a forecast of } \sigma_m^2 \text{ based on information } R_{m-1200}, \dots, R_{m-1}, \hat{\psi}, \quad (22)$$

for  $m=N+1, \dots, N+M$  and  $\hat{\psi}$  refers to the maximum likelihood estimate of the parameter vector  $\psi$ . In the case of realised volatility models, the conditioning also refers to the intraday returns associated with trading days  $m=N+1, \dots, N+M$ . The range of the day index  $m$  reflects the period from 17 October 2001 to 14 November 2003. The forecasts are based on a “rolling window” procedure: in each case, the sample is rolled forward by one trading day, while the size of the sample is kept constant at 1200.

The volatility is not observable, and as a result, the forecast error is also not observable. It is argued earlier that realised volatility (5), denoted by  $\tilde{\sigma}_m^2$ , is an accurate estimator, and therefore,  $\tilde{\sigma}_m^2$  is the variable to forecast for  $m=N+1, \dots, N+M$ . The volatility forecast obtained from model  $\mathcal{M}_k$  is denoted by  $\hat{\sigma}_{k,m}^2$  and is defined as in Eq. (22). It is difficult to determine whether, say, model  $\mathcal{M}_1$  outperforms model  $\mathcal{M}_2$ . This is a subjective matter that depends on the choice of a criterion. Various forecasting criteria can be considered to assess the predictive accuracy of a volatility model. In this paper, we use the following accuracy statistics or loss functions for a particular model  $\mathcal{M}_k$ :

The squared error	$L_{1,k,m} = (\hat{\sigma}_m^2 - \hat{\sigma}_{k,m}^2)^2$
The absolute error	$L_{2,k,m} =  \hat{\sigma}_m^2 - \hat{\sigma}_{k,m}^2 $
The squared error adjusted for heteroskedasticity	$L_{3,k,m} = (1 - \tilde{\sigma}_{k,m}^{-2} \hat{\sigma}_m^2)^2$
The absolute error adjusted for heteroskedasticity	$L_{4,k,m} =  1 - \tilde{\sigma}_{k,m}^{-2} \hat{\sigma}_m^2 $

Some of these loss functions are also considered by Andersen et al. (1999), Martens (2002) and Hansen and Lunde (2005).

When a particular loss function is smaller for model  $\mathcal{M}_1$  than for model  $\mathcal{M}_2$ , we can clearly not conclude that the forecasting performance of model  $\mathcal{M}_1$  is superior to that of

model  $\mathcal{M}_2$ . Such a conclusion cannot be made on the basis of just one criterion and just one sample. Recent work has focused on a testing framework for determining whether a particular model is outperformed by another model; see, for example, [Diebold and Mariano \(1995\)](#), [West \(1996\)](#) and [White \(2000\)](#). A further development of the White framework is known as the superior predictive ability (SPA) test and is proposed by [Hansen \(2001\)](#) where it is also shown that SPA has good power properties and is robust. We adopt the SPA test to investigate the relative performance of various volatility models. A similar study is carried out by [Hansen and Lunde \(2005\)](#) for a range of daily volatility models, mainly variants of GARCH models. In this study, we focus on the comparison of the forecast abilities of models based on high-frequency measures of volatility and models based on daily returns.

We consider  $K+1$  different models  $\mathcal{M}_k$  for  $k=0,1,\dots,K$  and which are discussed in Sections 3 and 4. For each model  $\mathcal{M}_k$ , we generate  $M$  volatility forecasts  $\hat{\sigma}_{m,k}^2$  for  $m=N+1,\dots,N+M$ . For every forecast, we calculate the loss function  $L_{i,m,k}$  as given above with  $i=1,\dots,4$ . A particular model  $\mathcal{M}_0$  is taken as the benchmark model. The loss function relative to the benchmark model is defined as

$$X_{k,m} \equiv L_{i,0,m} - L_{i,k,m}, \quad (23)$$

for some choice of loss function  $i$ . The unobserved actual volatility  $\sigma_m^2$  is the object of forecasting, while realised volatility  $\tilde{\sigma}_m^2$  is taken as its proxy for assessing the relative forecast accuracy of models. [Hansen and Lunde \(2003\)](#) argue that the use of a proxy may lead to inconsistencies in the comparison of models with respect to forecast accuracy. Consistency requires moment conditions and dependence structures for relative forecast losses  $X_{k,m}$ . These conditions are weak and likely to be fulfilled for at least  $L_{1,k,m}$  and  $L_{2,k,m}$ . Further, the results in [Hansen and Lunde \(2003\)](#) imply that the inconsistencies are proportionate with the precision of the proxy. The results in [Barndorff-Nielsen and Shephard \(2002\)](#) support the fact that realised volatility is a precise estimator of actual volatility. Assuming that models can be ranked consistently,  $\lambda_k \equiv E(X_{k,m})$  is well defined. When the benchmark or base model  $\mathcal{M}_0$  outperforms all other models, we have  $\lambda_k < 0$  for all models  $k=1,\dots,K$ . Therefore, the base model is not outperformed when it accepts the null hypothesis

$$\max_{k=1,\dots,K} \lambda_k \leq 0. \quad (24)$$

The associated test statistic proposed by [Hansen \(2001\)](#) is given by

$$T = \max_{k=1,\dots,K} \frac{\sqrt{M} \tilde{X}_k}{\hat{\omega}_{kk}} \quad (25)$$

with  $\hat{\omega}_{kk}^2$  as a consistent estimate of  $\omega_{kk}^2$ , and where

$$\tilde{X}_k = \frac{1}{M} \sum_{m=N+1}^{N+M} X_{k,m}, \quad \omega_{kk}^2 = \lim_{M \rightarrow \infty} \text{var}(\sqrt{M} \tilde{X}_k). \quad (26)$$

A consistent estimator of  $\omega_{kk}$  and the  $p$ -value of test statistic  $T$  can be obtained via a bootstrap procedure that is outlined below.

We apply the stationary bootstrap of [Politis and Romano \(1994\)](#) to compute the test statistic  $T$  and to assess its distributional properties. This procedure consists of building

new samples for  $X_{k,m}$  of length  $M$  which are constructed by randomly choosing subsamples of different lengths and putting these together. The lengths of the subsamples are independent and are drawn from a geometric distribution with mean  $q$ . The random lengths are ideally small but sufficiently large to reflect the serial dependence in the  $X_{k,m}$  series. After the inspection of correlograms of  $X_{k,m}$  for different models  $\mathcal{M}_k$ , we have taken  $q=0.5$ . The resulting bootstrap samples for  $X_{k,m}$  are denoted by  $X_{k,m}^i$ , and we consider  $B$  bootstraps,  $i=1, \dots, B$ . The empirical distribution implied by the bootstrap samples is taken as the distribution that has generated  $X_{k,m}$ . For each bootstrap the sample mean is computed as

$$\bar{X}_k^i = \frac{1}{M} \sum_{m=1}^M X_{k,m}^i, \quad i = 1, \dots, B. \quad (27)$$

The variation within these bootstrapped means is taken as the estimator of  $\omega_{kk}$ ; that is

$$\hat{\omega}_{kk} = \frac{1}{B} \sum_{i=1}^B \left( \bar{X}_k^i - \bar{\bar{X}}_k \right)^2, \quad \bar{\bar{X}}_k = \frac{1}{B} \sum_{i=1}^B \bar{X}_k^i. \quad (28)$$

The bootstrap estimate of  $\omega_{kk}$  allows the computation of the test statistic  $T$  in Eq. (25). Finally, the distribution of  $T$  under the null hypothesis can be empirically identified as follows. Define  $\bar{Z}_k^i$  as

$$\bar{Z}_k^i = \left( \bar{X}_k^i - \bar{\bar{X}}_k \right) \times 1_{\{\bar{\bar{X}}_k > -A_k\}}, \quad (29)$$

where  $A_k = \frac{1}{4} M^{-4} \hat{\omega}_{kk}$  and  $1_{\{\cdot\}}$  is an indicator function. The empirical distribution of

$$T^i = \max_{k=1, \dots, K} \frac{\sqrt{M} \bar{Z}_k^i}{\hat{\omega}_{kk}}, \quad (30)$$

converges to the distribution of  $T$  under the null hypothesis; see Hansen (2001) for more details. It follows directly that the  $p$ -value of  $T$  can be empirically identified as

$$\frac{1}{B} \sum_{i=1}^B 1_{\{T^i > T\}}. \quad (31)$$

More details of this procedure are detailed in Hansen (2001) and Hansen and Lunde (2005).

## 6. Empirical results

### 6.1. Parameter estimation

The estimation results obtained from realised and historical volatility models, as described in Sections 3 and 4, respectively, are presented in Table 2.<sup>9</sup> The results are

<sup>9</sup> Due to space limitations, we only present the empirical results of interest for this paper. The estimation results of other model specifications can be obtained from the authors upon request.

Table 2

Estimation results of realised and historical volatility using the standard & poor 100 stock index data set for the period 6 January 1997 to 15 November 2003

Model	UC ARMA		ARFIMA			Stochastic volatility model				GARCH model		
	UC-RV1	UC-RV2	In levels	In logs		With RV		With IV		With RV		With IV
$\mu$	0.955	0.953	0.813	−0.797	$\sigma^{*2}$	1.473	1.775	0.040	$\alpha$	0.102	−0.047	0.016
	0.058	0.224	0.779	1.434		0.180	0.166	0.024		0.010	0.019	0.017
$\phi_1$	0.451	0.422	0.051	−0.109	$\phi$	0.967	0.964	0.934	$\beta$	0.861	0.448	0.252
	0.025	0.130	0.040	0.031		0.011	0.018	0.029		0.016	0.070	0.103
$\phi_2$		0.976			$\gamma$		0.279	1.127	$\gamma$		0.908	0.126
		0.009					0.053	0.181			0.113	0.018
$d$			0.408	0.477								
			0.031	0.020								
$\sigma_{\eta_1}^2$	1.103	0.030			$\sigma_{\eta}^2$	0.026	0.014	0.028	$\omega$	0.075	0.312	−2.000
	0.040	0.012				0.008	0.007	0.012		0.017	0.065	0.284
$\sigma_{\eta_2}^2$		0.439										
		0.079										
$\sigma_u^2$	1.002	1.003	1.069	0.489								
$\theta_1$	0.259	0.257										
$\theta_2$		0.268										
$\ln L$	−2616.3	−2506.4	−2506.1	−1834.86		−2876.3	−2861.2	−2856.2		−2895.1	−2833.5	−2830.5
AIC	5238.6	5022.7	5018.3	3675.72		5758.2	5730.4	5720.4		5796.3	5675.0	5669.0
$Q(12)$	14.14	14.02	12.28	20.01		22.05	22.19	20.61		17.36	20.11	18.14

The realised volatility model UC-RVJ (with  $J=1,2$ ) is for  $\tilde{\sigma}_n^2$  and given by Eq. (12). The long memory ARFIMA model is also for  $\tilde{\sigma}_n^2$  (in levels and logs) and is given by Eq. (13). The stochastic volatility model is for  $R_n$  and is given by Eq. (14), together with Eq. (16) when an explanatory variable is included in the log-volatility process. The GARCH model is also for  $R_n$  and is given by Eq. (18), together with Eq. (20) when an explanatory variable is included in the log-volatility process. Parameter estimates are reported together with asymptotic 95% standard errors, which are obtained using the delta method since some coefficients are transformed for estimation. The Akaike information criterion (AIC) is calculated as  $-2(\ln L)+2p$  where  $p$  is the number of coefficients that is estimated. The Box–Ljung portmanteau statistic  $Q(l)$  is for the observation errors and is asymptotically  $\chi^2$  distributed with  $l$  degrees of freedom.

based on observations from the full sample period 6 January 1997 to 14 November 2003. All models indicate high persistency in volatility dynamics as it is expected from stock index prices. Unobserved ARMA component models (UC-RV) are estimated with one (UC-RV1) and two (UC-RV2) factors. The coefficients  $\mu$ ,  $\phi_1$ ,  $\phi_2$ ,  $\sigma_{\eta 1}^2$  and  $\sigma_{\eta 2}^2$  are estimated, and standard errors are provided, while the coefficients  $\sigma_u^2$ ,  $\theta_1$  and  $\theta_2$  are functions of the estimated coefficients, as is apparent from Eqs. (10) and (11). The ARMA factor in UC-RV1 with autoregressive coefficient 0.45 and moving average coefficient 0.26 does not imply the persistent dynamic structure, as indicated by the sample correlogram of realised volatility  $\hat{\sigma}_n^2$  given in Fig. 1. However, a higher level of volatility persistency is obtained when two ARMA factors are considered. The second factor in UC-RV2 has a higher persistency with autoregressive coefficient 0.98. The log-likelihood function is larger for the two-factor UC-RV2 model than for the one-factor UC-RV1 model, which is confirmed by a lower Akaike information criterion (AIC) value for UC-RV2. More ARMA factors have also been considered, but two factors have proved to be sufficient for this data set. Both UC-RV models are successful in capturing the dynamics in the series, given the relatively small values of the Box–Ljung  $Q$  statistics.

The fractional integration parameter of the ARFIMA model is estimated as 0.408 for realised volatility in levels and as 0.478 for logged realised volatility. In the latter case, the  $d$  estimate exceeds the estimate found by Andersen et al. (2003) who report the value 0.441. It should be noted that this and related studies employ different ARFIMA model specifications, different estimation methods and different data sets.<sup>10</sup> Further, in the case of logged realised volatility, the process may not be covariance-stationary, as the estimate of  $d$  is close to the boundary value of 0.5. Some evidence of nonstationarity is also found in the high persistency implied by the second ARMA component in the UC-RV2 model. From these in-sample results, we may conclude that a parsimonious and effective description of the dynamics in the S&P 100 realised volatility is provided by the ARFIMA(1, $d$ ,0) model. The dynamics of realised volatility seem to be best captured by the ARFIMA model since it has the lowest Box–Ljung test statistic. The ARFIMA log-likelihood value and AIC for logged realised volatility cannot be compared with the other realised volatility models since it refers to different values of the dependent variable. The Box–Ljung statistic is the highest for the log ARFIMA model, but this is probably due to the fact that logged series have usually less atypical observations (outliers). Aberrant observations in a time series have a destructive effect on the serial correlation structure.

Volatility persistency for the historical SV and GARCH models are estimated as statistically significant and close to unity with  $\phi=0.967$  and  $\hat{\alpha}+\hat{\beta}$  equal to 0.963, which confirms earlier findings in the literature with regard to daily stock index return series. The estimated SV model with a higher log-likelihood value is preferred since it indicates a better fit compared to the GARCH model. However, one should be careful in comparing

<sup>10</sup> Andersen et al. (2003) do not estimate the fractional parameter but obtain the  $d$  value from a prior analysis and fix it for the estimation of the other parameters in their vector autoregressive fractional model for three exchange rates.

likelihood values since the involved computations for SV and GARCH models are very different, and models are nonnested.<sup>11</sup>

The inclusion of the realised volatility series in the volatility equation of the SV model has a significant effect on the fit of the model as is shown by a decrease of 27.8 in the AIC value and by the significance of the estimated regression coefficient  $\gamma$ . The estimate of the persistence parameter  $\phi$  has not changed significantly. The change in the estimated values for  $\sigma^2$  and  $\sigma_\eta^2$  can be explained by the necessary rescaling due to the inclusion of the regression variable  $\ln \hat{\sigma}_\eta^2$ . Similar conclusions can be made for the inclusion of implied volatility in logs; that is  $\ln s_n^2$ , in the log-volatility equation of the SV model. However, the effect is even stronger since the AIC decreases by 37.8 in relation to the standard SV model. Also, the regression coefficient is more significant compared to the significance of  $\gamma$  in the realised volatility case. Therefore, the SV model with implied volatility in the volatility equation may be preferred based on these in-sample results.

The estimation results for the GARCH model are typical for stock index series and can be found elsewhere in the literature. It is noticeable that within the GARCH framework, the inclusion of realised or implied volatility as explanatory variables for volatility has a large impact, larger than in the SV case. The AIC value decreases by 61.6 and 64.6 for the two volatilities. Also, in both cases, the regression coefficients are significant. However, the inclusion of these regressors diminishes the persistency of the volatility dynamics. The estimated coefficient  $\beta$  remains significant in both cases. In the case where implied volatility is included, the estimated  $\alpha$  coefficient is not significant. This case implies that the daily (squared) returns do not play a role in the estimation of  $\sigma_n^2$  which in fact only depends on past values of implied volatilities; that is

$$\hat{\sigma}_n^2 = \hat{\gamma} \sum_{j=1}^{\infty} \hat{\beta}^{j-1} s_{n-j}^2.$$

Here,  $\beta$  is estimated as  $\hat{\beta}=0.25$ , implying that effectively, only  $s_{n-1}^2$  and, to a smaller extent,  $s_{n-2}^2$  determine  $\hat{\sigma}_n^2$ .

## 6.2. Some forecasting results

Volatility forecasts are constructed for realised and historical volatility models, as described in Section 5. The evaluation period ranges from 17 October 2001 to 14 November 2003. In this period, for each trading day  $n$ , a volatility forecast is generated for each considered model which is estimated using information from the last 1200 days ( $n-1, \dots, n-1200$ ). The one-step ahead forecasts for the evaluation period are presented in Fig. 2 together with realised volatility (displayed as dots). Some interesting aspects of the forecast results can be observed since in this sample, a small group of days displays high volatility while most other days have low volatility. The GARCH and SV forecasts

<sup>11</sup> The estimation results for the GARCH model confirm the empirical results of Blair et al. (2001) who examine Standard & Poor's 100 stock index returns over the earlier 1987 to 1992 period and find values for  $\gamma$  similar to ours. In contrast, Martens (2002) reports on much smaller and statistically insignificant  $\gamma$  estimates for returns on Standard & Poor's 500 futures; see also Taylor and Xu (1997).

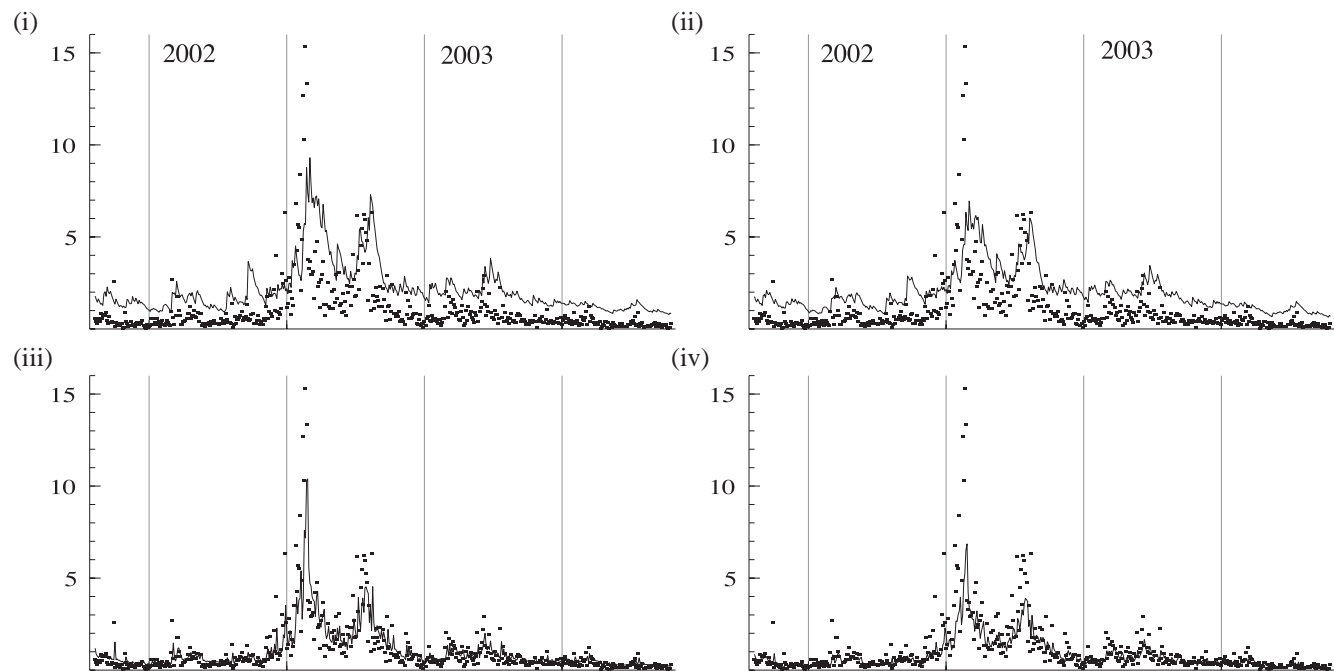


Fig. 2. One-day ahead volatility forecasts from (i) GARCH, (ii) SV, (iii) UC-RV2 and (iv) log ARFIMA-RV models for the period between 17 October 2001 and 14 November 2003 (525 days).



Table 3

Out-of-sample forecasting criteria (evaluated against  $\hat{\sigma}_m^2$ ) for Standard & Poor's 100 evaluation period 17 October 2001 to 14 November 2003

Model $\mathcal{M}_k$	Forecast loss functions				
	MSE	MAE	HMSE	HMAE	$R^2$
UC-RV1	1.248	0.613	2.495	1.240	0.522
UC-RV2	0.996	0.505	1.546	0.792	0.593
ARFIMA-RV	0.991	0.508	1.610	0.813	0.598
ARFIMA-RV (log)	1.149	0.472	1.030	0.555	0.597
SV	2.433	1.240	5.080	2.948	0.386
SV RV	2.256	1.037	3.368	2.063	0.437
SV IV	3.132	1.082	3.422	2.048	0.343
GARCH	2.837	1.348	5.339	3.174	0.405
GARCH RV	3.134	1.228	4.603	2.738	0.605
GARCH IV	2.872	1.297	5.079	2.720	0.419

The MSE for model  $\mathcal{M}_k$  is defined as  $M^{-1} \sum_{m=N+1}^{N+M} L_{1,k,m}$ . In a similar way, MAE (based on  $L_{2,k,m}$ ), HMSE (based on  $L_{3,k,m}$ ) and HMAE (based on  $L_{4,k,m}$ ) are defined. The goodness-of-fit  $R^2$  statistic is based on the OLS regression applied to  $\sigma_m = a + b\hat{\sigma}_m + u_m$  where  $a$  and  $b$  are regression coefficients and with realised volatility  $\hat{\sigma}_m$ , forecast  $\hat{\sigma}_m$  as defined in Eq. (22) and error  $u_m$  for  $m=N+1, \dots, N+M$  and  $M=525$ . In a similar forecasting context, these so-called Mincer–Zarnowitz regressions are also carried out by Andersen et al. (2003).

are clearly biased when volatility is low. The GARCH forecasts are also too high for days when volatility is relatively high. The positive and persistent forecast bias may be explained from the fact that the GARCH and SV forecasts are extracted from daily returns that are subject to measurement noise. The volatility estimates are not corrected for this measurement variance, and therefore, the forecast bias is positive. In contrast, the realised volatility models appear to produce forecasts without positive forecast bias. The log ARFIMA-RV forecasts are clearly the least sensitive to jumps in volatility.

Table 3 presents the forecast accuracy statistics that consist of the means of four loss functions:  $L_{1,k,m}$ , mean squared error (MSE);  $L_{2,k,m}$ , mean absolute error (MAE);  $L_{3,k,m}$ , heteroskedastic adjusted MSE (HMSE);  $L_{4,k,m}$ , heteroskedastic adjusted MAE (HMAE).<sup>12</sup> In terms of MSE, the UC-RV2 and ARFIMA-RV models produce the smallest values for this sample. In terms of the other criteria, the log ARFIMA-RV and UC-RV2 models produce the smallest values. The heteroskedastic corrected versions of MSE and MAE give relatively less weight to forecast errors associated with high values of realised volatility, while the MSE give relatively more weight to these errors. Therefore, both of these statistics are of interest on their own merits. In case of value-at-risk applications, one may be more interested in the accurate forecasting of high volatility rather than low volatility. This implies that the MSE criterion is the relevant loss function for applications in risk management.

Realised volatility models provide more accurate forecasts compared to the historical SV and GARCH models that are based on daily returns. Within the group of historical models, the SV model with either realised or implied volatility as an explanatory variable for

<sup>12</sup> For example, we define the MSE for model  $\mathcal{M}_k$  by  $M^{-1} \sum_{m=N+1}^{N+M} L_{1,k,m}$ .

volatility is providing the most accurate forecasts. This holds when considering the different loss functions. The accuracy of GARCH forecasts varies, but overall, their forecasts are less accurate. Within the group of realised volatility models, there is a clear distinction in accuracy. For the MSE loss function, the ARFIMA-RV and UC-RV2 models produce the smallest values, while for the other loss functions, the log ARFIMA-RV model clearly outperforms the other models in forecasting accuracy.

For completeness, we also report the goodness-of-fit measure  $R^2$  for regressing  $\tilde{\sigma}_m$  on a constant and the volatility forecast  $\hat{\sigma}_m$  for  $m=N+1, \dots, N+M$ . This so-called Mincer–Zarnowitz regression is used in many volatility forecast studies. The goodness-of-fit statistics confirm the earlier conclusions, although the GARCH and SV models produce relatively high  $R^2$  values, especially the GARCH model with realised volatility. This can be explained from the fact that the  $R^2$  measure corrects for the bias in forecast errors since it compares the total fit of the Mincer–Zarnowitz regression with the sum of squared mean deviations of forecast errors. Therefore, it shows that the pattern of one-step ahead forecasts obtained from the GARCH model with realised volatility is close to the pattern of volatility itself. The realised volatility models, however, produce overall the highest  $R^2$  values. These results confirm similar findings reported in Andersen et al. (2003) for exchange rate data.

Finally, to get some insight in how forecasts evolve over time in our study, in Fig. 3, we have zoomed in on the one-step ahead forecasts between 9 September 2002 and 18 November 2002 (51 trading days). The forecasts of GARCH and SV respond slowly to the changes from high to low and from low to high volatility. The forecasts of these models are obtained from weighted averages of past squared returns with relatively long exponentially declining weight patterns. This may explain the slow response of the forecasts to changes in volatility, and, together with the volatility measurement error in daily squared returns, it may provide an explanation for the relatively poor performance of historical models in volatility forecasting. The inclusion of realised volatility shows an improvement in both the GARCH and SV models. Especially in the latter case, it can be observed from Fig. 3 that the forecasts are more responsive to changes in volatility. The positive forecast bias is reduced when volatility is low, and it produces more accurate forecasts when volatility is high.

In the case of realised volatility models, it is shown that the more successful models in forecasting are the ones that produce a smoother series of forecasts (UC-RV2 and log ARFIMA-RV). It is surprising to notice that relative smooth forecast patterns are a virtue in forecasting volatility.

### 6.3. Superiority prediction tests

Although, the forecasting results have clearly shown that the ARFIMA model for logged realised volatility is the best model for prediction, and that the SV model with realised volatility as an explanatory variable in the volatility equation produces the best forecasts within the class of historical volatility models. However, these results only apply to one selected sample. To assess whether the same outcomes are obtained for similar samples, we need to apply the methodology of White (2000) and Hansen (2001), as discussed in Section 5. The computation of superior predictive ability (SPA)

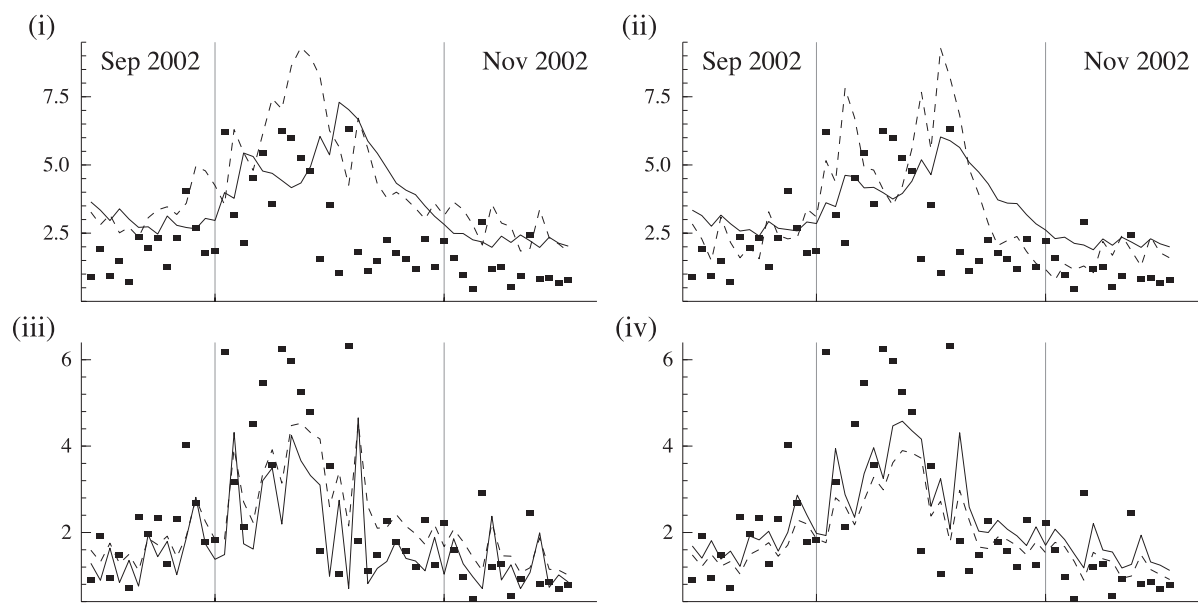


Fig. 3. Realised volatility (as dots) and one-day ahead volatility forecasts from (i) GARCH (solid) and GARCH with RV (dashed), (ii) SV (solid) and SV with RV (dashed), (iii) UC-RV1 (solid) and UC-RV2 (dashed) and (iv) ARFIMA-RV (solid) and log ARFIMA-RV (dashed) models for the period between 9 September 2002 and 18 November 2002 (day 225 to 275).

test (25) requires the stationary bootstrap of loss series  $L_{i,k,m}$  over the sample  $m=N+1, \dots, N+M$  for a specific loss function  $i$  and for a range of models  $\mathcal{M}_k$  with  $k=0, \dots, K$ . Computing  $\bar{X}_k$  in Eq. (26) is based on the actual forecasts, while computing  $\hat{\omega}_{kk}$  in Eq. (28) requires the bootstraps. The computation of the  $p$ -values as in Eq. (31) also requires the bootstraps.

Tables 4 and 5 present the SPA results for the two classes of volatility models. The individual  $\bar{X}_k$  is reported for each model that is considered as a base model  $\mathcal{M}_0$ , while the remaining models are treated as rival models  $\mathcal{M}_k$ . All considered models in this study are consecutively taken as base models. Further, the various statistics are reported for the four different loss functions. As a result, and in the way the  $\bar{X}_k$  statistics are presented, the tables (apart from the last column) consist of four blocks of skewed symmetric matrices. The last columns of Tables 4 and 5 report the  $p$ -values of the SPA tests for each considered base model. A high  $p$ -value means that the null hypothesis (24) of “base model is not outperformed” cannot be rejected. The earlier forecast results have seen a large difference in forecast accuracy of historical and realised volatility models. Therefore, the SPA results are presented separately for these two classes of models.

In Table 4, the SPA results for the realised volatility models are presented. The average of daily squared forecast error differences between UC-RV1 and UC-RV2 models is 2.21, which means that the squared forecast error of UC-RV1 was on average higher than of UC-RV2. The average values for UC-RV1 in relation to the rival ARFIMA and log ARFIMA models are 2.39 and 1.15, respectively. This result is not good for UC-RV1 as is

Table 4

Superior predictive ability (SPA) test  $\lambda_k$  for realised volatility models and the Standard & Poor’s 100 evaluation period 17 October 2001 to 14 November 2003

Loss	Base model	Alternative models				$p$ -value $T^*$
		UC-RV1	UC-RV2	ARFIMA	ARF.log	
MSE	UC-RV1	–	2.21	2.39	1.15	0.028
	UC-RV2	–2.21	–	0.38	–0.99	0.564
	ARFIMA	–2.39	–0.38	–	–1.07	1 <sup>a</sup>
	ARFIMA log	–1.15	0.99	1.07	–	0.197
MAE	UC-RV1	–	7.05	7.28	9.08	0.000
	UC-RV2	–7.05	–	–0.79	1.58	0.053
	ARFIMA	–7.28	0.79	–	2.20	0.019
	ARFIMA log	–9.08	–1.85	–2.20	–	1 <sup>a</sup>
HMSE	UC-RV1	–	2.80	2.97	2.89	0.011
	UC-RV2	–2.80	–	–1.21	3.05	0.007
	ARFIMA	–2.97	1.21	–	2.63	0.023
	ARFIMA log	–2.89	–3.05	–2.63	–	1 <sup>a</sup>
HMAE	UC-RV1	–	7.96	7.82	8.78	0.000
	UC-RV2	–7.96	–	2.64	8.84	0.000
	ARFIMA	–7.82	2.64	–	8.64	0.000
	ARFIMA log	–8.78	–8.84	–8.64	–	1 <sup>a</sup>

A positive value in column of an alternative model  $\mathcal{M}_k$  indicates that it is superior to the base model according to a specific loss function. Final column has  $p$ -value of  $T^*$  test and can be interpreted as the intensity of the base model producing superior forecast.

<sup>a</sup> A unity value indicates here that the test is not computed since the model outperformed all others.

Table 5

Superior predictive ability (SPA) test  $\lambda_k$  for daily SV models and the Standard & Poor's 100 evaluation period 17 October 2001 to 14 November 2003

Loss	Base model	Alternative models						<i>p</i> -value $T^*$
		SV	SV RV	SV IV	GARCH	GA.RV	GA.IV	
MSE	SV	–	0.73	–1.26	–2.42	–1.02	–1.91	0.0604
	SV RV	–0.73	–	1.68	–3.72	–1.34	–2.06	–1 <sup>a</sup>
	SV IV	1.26	1.68	–	0.60	–0.00	0.44	0.089
	GARCH	2.42	3.72	–0.60	–	–0.45	–0.13	0.002
	GARCH RV	1.02	1.34	0.00	0.45	–	0.37	0.164
	GARCH IV	1.91	2.06	–0.44	0.13	–0.37	–	0.027
MAE	SV	–	5.86	2.42	–5.32	0.21	–1.02	0.000
	SV RV	–5.86	–	–0.76	–10.5	–3.41	–4.05	1 <sup>a</sup>
	SV IV	–2.42	0.76	–	–4.50	–1.79	–2.69	0.400
	GARCH	5.32	10.5	4.50	–	2.05	0.80	0.000
	GARCH RV	–0.21	3.41	1.79	–2.05	–	–0.89	0.004
	GARCH IV	1.02	4.05	2.69	–0.80	0.89	–	0.000
HMSE	SV	–	2.01	1.94	–6.00	0.94	0.04	0.070
	SV RV	–2.01	–	–0.55	–2.32	–3.07	–1.90	1a
	SV IV	–1.94	0.55	–	2.26	–3.06	–1.83	0.575
	GARCH	6.00	2.32	2.26	–	1.45	1.15	0.000
	GARCH RV	–0.94	3.07	3.06	–1.46	–	–0.80	0.009
	GARCH IV	–0.04	1.90	1.83	–1.15	0.80	–	0.089
HMAE	SV	–	7.76	7.99	–10.1	2.26	1.77	0.000
	SV RV	–7.76	–	0.25	–9.11	–5.46	–4.39	0.405
	SV IV	–7.99	–0.25	–	–9.66	–6.69	–4.48	1 <sup>a</sup>
	GARCH	10.1	9.11	9.66	–	4.77	3.17	0.000
	GARCH RV	–2.26	5.46	6.69	–4.77	–	0.11	0.000
	GARCH IV	–1.77	4.39	4.48	–3.17	–0.11	–	0.000

A positive value in column of an alternative model  $\mathcal{M}_k$  indicates it is superior to the base model according to a specific loss function. Final column has *p*-value of  $T^*$  test and can be interpreted as the intensity of the base model producing superior forecast.

<sup>a</sup> A unity value indicates here that model outperformed others.

also reflected by its *p*-value of zero for the SPA test. Overall, Table 4 shows that, for all loss functions, the UC-RV2 and ARFIMA models are very close in their good forecasting performance. However, the log ARFIMA model is clearly the preferred model when we rely on the loss functions  $L_2$ ,  $L_3$  and  $L_4$ . This is particularly noticeable when we consider the *p*-value of the SPA test. For the last three loss functions, its *p*-value is unity, implying it is superior to all other models that are considered. The first loss function (squared forecast error), however, shows that ARFIMA and UC-RV2 forecasts are to be preferred which confirms the earlier findings from Table 3.

Finally, in Table 5 the results are even more dramatic in its clarity. For all loss functions, it can be concluded that forecasts from the SV class of models are to be preferred. Overall, the SV model with logged realised volatility as an explanatory variable for log-volatility is the outstanding model within its class.

These forecasting results show the importance of realised volatility as a predictor of future volatility. We confirm the findings of Christensen and Prabhala (1998), Fleming

(1998), Blair et al. (2001) and Giot (2003) that forecasts of volatility extracted from models for daily returns are less accurate than forecasts based on implied volatility. However, in this study, we have found that models based on realised volatility outperform models with implied volatility when forecasting the volatility in the S&P 100 index series. The importance of realised volatility for forecasting is also found by Andersen et al. (2003), although the empirical part of their study is based on various GARCH and (trivariate) ARFIMA models for exchange rates while this study considers univariate ARFIMA models together with unobserved component models (based on Lévy-driven OU processes) for a stock index time series.

## 7. Summary and conclusions

In this paper, we examine the predictive abilities of four classes of volatility models by comparing the forecasts with realised volatility measures that are defined as scaled sums of squared intraday returns. Time series of realised volatility can be modelled by unobserved components (UC-RV) and by autoregressive fractionally integrated moving average (ARFIMA-RV) models. Forecasts can be generated from these models after they are estimated using relatively straightforward methods. Daily returns can be modelled by stochastic volatility (SV) and generalised autoregressive conditional heteroskedasticity (GARCH) models from which volatility estimates and forecasts can be obtained. In the case of GARCH models, estimation and forecasting belong to the standard toolbox of many applied workers. More involved statistical procedures are required for estimating and forecasting volatility on the basis of SV models. We empirically investigate the one-step ahead forecasting performance of the various models for the Standard & Poor's 100 stock index over the period 17 October 2001 to 14 November 2003. The most accurate forecasts are obtained with the log ARFIMA-RV model and closely followed by the UC-RV model. Within the classes of historical volatility models, the SV model with realised volatility incorporated as an explanatory variable in the volatility equation is best. These conclusions are not only based on a standard forecasting accuracy assessment using mean squared forecast error and mean absolute forecast error statistics, formal prediction superiority tests are also carried out. The implementation closely follows the developments reported in White (2000) and Hansen (2001). This extensive study may have shed some light on the discussion on which model is mostly suited for the forecasting of volatility. In this respect, we would like to stress the importance of realised volatility in empirical work. Recent work by Andersen et al. (2001a), Barndorff-Nielsen and Shephard (2001, 2002), Andersen et al. (2003) and work in progress on realised volatility is therefore highly relevant for the forecasting of volatility.

## Acknowledgements

We would like to thank Marius Ooms and two referees for their comments, which have led to a much improved paper.

## References

- Andersen, T.G., Bollerslev, T., 1997. Intraday seasonality and volatility persistence in foreign exchange and equity markets. *Journal of Empirical Finance* 4, 115–158.
- Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39, 885–905.
- Andersen, T.G., Bollerslev, T., Lange, S., 1999. Forecasting financial market volatility: sample frequency vis-à-vis forecast horizon. *Journal of Empirical Finance* 6, 457–477.
- Andersen, T.G., Bollerslev, T., Diebold, F.X., 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., Labys, P., 2001b. The distribution of exchange rate volatility. *Journal of the American Statistical Association* 96, 42–55.
- Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realised volatility. *Econometrica* 71, 579–625.
- Areal, N.M.P.C., Taylor, S.J., 2002. The realised volatility of FTSE-100 future prices. *Journal of Future Markets* 22, 627–648.
- Barndorff-Nielsen, O.E., Shephard, N., 2001. Non-Gaussian OU based models and some of their uses in financial economics (with discussion). *Journal of the Royal Statistical Society. Series B* 63, 167–241.
- Barndorff-Nielsen, O.E., Shephard, N., 2002. Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society. Series B* 64, 253–280.
- Barndorff-Nielsen, O.E., Shephard, N., 2004a. Measuring and forecasting financial variability using realised variance. In: Harvey, A.C., Koopman, S.J., Shephard, N. (Eds.), *State Space and Unobserved Components Models: Theory and Applications*. Cambridge University Press, Cambridge.
- Barndorff-Nielsen, O.E., Shephard, N., 2004b. Power and bipower variation with stochastic volatility and jumps (with discussion). *Journal of Financial Econometrics* 2, 1–37.
- Barucci, E., Reno, R., 2002. On measuring volatility of diffusion processes with high frequency data. *Economics Letters* 74, 371–378.
- Beckers, S., 1981. Standard deviations implied in options prices as predictors of future stock price variability. *Journal of Banking and Finance* 5, 363–381.
- Bera, A.K., Higgins, M.L., 1993. ARCH models: properties, estimation and testing. *Journal of Economic Surveys* 7, 305–366.
- Blair, B.J., Poon, S.H., Taylor, S.J., 2001. Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high frequency returns. *Journal of Econometrics* 105, 5–26.
- Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.
- Bollerslev, T., Chou, R.Y., Kroner, K.F., 1992. ARCH modeling in finance: a review of the theory and empirical evidence. *Journal of Econometrics* 52, 5–59.
- Bollerslev, T., Engle, R.F., Nelson, D.B., 1994. ARCH models. In: Engle, R.F., McFadden, D.L. (Eds.), *Handbook of Econometrics*, vol. 4. Elsevier, Amsterdam, pp. 2959–3038.
- Bos, C.S., Franses, P.H., Ooms, M., 2002. Inflation, forecast intervals and long memory regression models. *International Journal of Forecasting* 18, 243–264.
- Brodsky, J., Hurvich, C.M., 1999. Multi-step forecasting for long-memory process. *Journal of Forecasting* 18, 59–75.
- Chiras, D.P., Manaster, S., 1978. The information content of options prices and a test of market efficiency. *Journal of Financial Economics* 6, 213–234.
- Christensen, B.J., Prabhala, N.R., 1998. The relation between implied and realised volatility. *Journal of Financial Economics* 50, 125–150.
- Corsi, F., Zumbach, G., Müller, U., Dacorogna, M., 2001. Consistent high-precision volatility from high-frequency data. *Economic Notes* 30, 183–204.
- de Jong, P., Shephard, N., 1995. The simulation smoother for time series models. *Biometrika* 82, 339–350.
- Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13, 253–263.
- Doomik, J.A., 2001. *Ox: Object Oriented Matrix Programming*, 3.0. Timberlake Consultants Press, London.



- Doornik, J.A., Ooms, M., 2001. A Package for Estimating, Forecasting and Simulating Arfima Models: Arfima Package 1.01 for Ox, Working Paper. Nuffield College, Oxford.
- Doornik, J.A., Ooms, M., 2003. Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. *Computational Statistics & Data Analysis* 42, 333–348.
- Durbin, J., Koopman, S.J., 1997. Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika* 84, 669–684.
- Durbin, J., Koopman, S.J., 2000. Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives (with discussion). *Journal of the Royal Statistical Society. Series B* 62, 3–56.
- Durbin, J., Koopman, S.J., 2001. *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford.
- Durbin, J., Koopman, S.J., 2002. A simple and efficient simulation smoother for state space time series analysis. *Biometrika* 89, 603–616.
- Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica* 50, 987–1007.
- Feinstein, S.P. (1995). The Black–Scholes Formula is Nearly Linear in  $\sigma$  for At-The-Money Options; Therefore Implied Volatilities From At-The-Money Options are Virtually Unbiased. Working Paper, Boston University.
- Fleming, J., 1998. The quality of market volatility forecast implied by S&P 100 index option prices. *Journal of Empirical Finance* 5, 317–345.
- Fleming, J., Whaley, R.E., 1994. The value of wildcards options. *Journal of Finance* 49, 215–236.
- Fleming, J., Ostdiek, B., Whaley, R.E., 1995. Predicting stock market volatility: a new measure. *Journal of Futures Markets* 15, 265–302.
- French, K.R., Schwert, G.W., Stambaugh, R.F., 1987. Expected stock returns and volatility. *Journal of Financial Economics* 19, 3–29.
- Ghysels, E., Harvey, A.C., Renault, E., 1996. Stochastic volatility. In: Maddala, G.S., Rao, C.R. (Eds.), *Handbook of Statistics, Statistical Methods in Finance*, vol. 14. North-Holland, Amsterdam, pp. 119–191.
- Giot, P., 2003. The information content of implied volatility in agricultural commodity markets. *Journal of Futures Markets* 23, 441–454.
- Giot, P., Laurent, S., 2004. Modeling daily value-at-risk using realized volatility and ARCH type models. *Journal of Empirical Finance* 11, 379–398.
- Hansen, P.R., 2001. A test for superior predictive ability. Discussion Paper. Brown University Working Paper 2001-06.
- Hansen, P.R., Lunde, A., 2002. Volatility estimation using high frequency data with partial availability. Discussion Paper. Brown University Working Paper.
- Hansen, P.R., Lunde, A., 2003. Consistent preordering with an estimated criterion function, with an application to the evaluation and comparison of volatility models. Discussion Paper. Brown University Working paper 2003-01.
- Hansen, P.R., Lunde, A., 2005. A forecast comparison of volatility models: does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, forthcoming.
- Harvey, C.R., Whaley, R.E., 1992. Dividends and S&P 100 index option valuation. *Journal of Futures Markets* 12 (2), 123–137.
- Harvey, A.C., Ruiz, E., Shepard, N., 1994. Multivariate stochastic variance models. *Review of Economic Studies* 61, 247–264.
- Hull, J., White, A., 1987. The pricing of options on assets with stochastic volatilities. *Journal of Finance* 42, 281–300.
- Koopman, S.J., Shephard, N., Doornik, J., 1999. Statistical algorithms for models in state space using SsfPack 2.2. *Econometrics Journal* 2, 113–166.
- Latané, H.A., Rendleman, R.J., 1976. Standard deviations of stock price ratios implied in option prices. *Journal of Finance* 31, 369–381.
- Laurent, S., Peters, J.P., 2002. G@RCH 2.2: an Ox package for estimating and forecasting various ARCH models. *Journal of Economic Surveys* 16, 447–484.
- Lee, K.M., Koopman, S.J., 2004. Estimating stochastic volatility models: a comparison of two importance samplers. *Studies in Nonlinear Dynamics and Econometrics* 8 (Article 5).
- Malliavin, P., Mancino, M.E., 2002. Fourier series method for measurement of multivariate volatilities. *Finance and Stochastics* 6, 49–61.

- Martens, M., 2001. Forecasting daily exchange rate volatility using intraday returns. *Journal of International Money and Finance* 20, 1–23.
- Martens, M., 2002. Measuring and forecasting S&P 500 index-futures volatility using high-frequency data. *Journal of Futures Markets* 22, 497–518.
- Meddahi, N., 2003. ARMA representation of integrated and realized variances. *Econometrics Journal* 6, 335–356.
- Politis, D.N., Romano, J.P., 1994. The stationary bootstrap. *Journal of the American Statistical Association* 89, 1303–1313.
- Poon, S.H., Granger, C., 2003. Forecasting volatility in financial markets: a review. *Journal of Economic Literature* 41, 478–539.
- Poon, S.H., 1992. Stock returns and volatility: an empirical study of the UK stock market. *Journal of Banking and Finance* 16, 37–59.
- Ripley, B.D., 1987. *Stochastic Simulation*. Wiley, New York.
- Sandmann, G., Koopman, S.J., 1988. Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics* 87, 271–301.
- Shephard, N., 1996. Statistical aspects of ARCH and stochastic volatility. In: Cox, D.R., Hinkley, D.V., Barndorff-Nielsen, O.E. (Eds.), *Time Series Models in Econometrics, Finance and Other Fields*, Number 65 in *Monographs on Statistics and Applied Probability*. Chapman and Hall, London, pp. 1–67.
- Shephard, N., Pitt, M.K., 1997. Likelihood analysis of non-Gaussian measurement time series. *Biometrika* 84, 653–667.
- Sowell, F., 1992. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics* 53, 165–188.
- Taylor, S.J., 1986. *Modeling Financial Time Series*. John Wiley and Sons, Chichester.
- Taylor, S.J., Xu, X., 1997. The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance* 4, 317–340.
- West, K.D., 1996. Asymptotic inference about predictive ability. *Econometrica* 64, 1067–1084.
- White, H., 2000. A reality check for data snooping. *Econometrica* 68, 1097–1126.