Journal of Econometrics 147 (2008) 17–33

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Out of sample forecasts of quadratic variation☆

Yacine Aït-Sahalia[a,b,*], Loriano Mancini[c]

[a] Department of Economics, Princeton University, United States
[b] NBER, United States
[c] Swiss Banking Institute, University of Zurich, Switzerland

ARTICLE INFO

Article history:
Available online 18 September 2008

JEL classification:
C14
C22
C53

Keywords:
Market microstructure noise
High frequency data
Measurement error
Realized volatility
Two scales realized volatility
Out of sample forecasts

ABSTRACT

We compare the forecasts of Quadratic Variation given by the Realized Volatility (RV) and the Two Scales Realized Volatility (TSRV) computed from high frequency data in the presence of market microstructure noise, under several different dynamics for the volatility process and assumptions on the noise. We show that TSRV largely outperforms RV, whether looking at bias, variance, RMSE or out-of-sample forecasting ability. An empirical application to all DJIA stocks confirms the simulation results.

1. Introduction

In financial econometrics, one is often interested in estimating the volatility of the efficient log-price process

$$dX_t = \mu_t dt + \sigma_t dW_t \tag{1.1}$$

using discretely sampled data on the transaction price process at times $0, \Delta, \ldots, n\Delta = T$. The volatility of asset returns plays an important role in derivative pricing and hedging, asset allocation, risk management—pretty much all of asset pricing. When σ_t is itself a stochastic process, a main object of interest is the quadratic variation or integrated variance (IV)

$$\langle X, X\rangle_T = \int_0^T \sigma_t^2 dt \tag{1.2}$$

over a fixed time period $[0, T]$, say one day. The usual estimator of $\langle X, X\rangle_T$ is the realized volatility (RV), which is simply the sum of observed squared log-returns

$$[X, X]_T = \sum_{i=1}^{n} (X_{t_{i+1}} - X_{t_i})^2 \tag{1.3}$$

and this estimator has been used extensively in the recent literature: see e.g., Andersen and Bollerslev (1998) and Barndorff-Nielsen and Shephard (2002).

In theory, sampling at increasingly higher frequency should deliver, in the limit, a consistent estimator of the quadratic variation. The sum $[X, X]_T$ consistently estimates the integral $\langle X, X\rangle_T$ as has been well-known in the stochastic processes literature. Further, the sum converges to the integral with a known distribution, a result dating back to Jacod (1994) and Jacod and Protter (1998). Selecting Δ as small as possible ($=n$ as large as possible) is optimal.

Everything would be perfect if it were not for the fact that financial asset returns are subject, especially at high frequency, to a vast array of frictions. As a result, a perhaps more realistic model is one where the observed transaction log-price is not X but rather Y, the sum of an unobservable efficient price X and a noise component due to the imperfections of the trading process, ε:

$$Y_t = X_t + \varepsilon_t. \tag{1.4}$$

☆ Financial support from the NSF under grants SES-0350772 and DMS-0532370 (Aït-Sahalia) and from the University Research Priority Program "Finance and Financial Markets" University of Zurich and the NCCR-FinRisk Swiss National Science Foundation (Mancini) is gratefully acknowledged. For helpful comments, we thank two anonymous referees, Patrick Cheridito, Fulvio Corsi and seminar participants at the conferences "Risk Measures & Risk Management for High-Frequency Data", EURANDOM, 2006, "Microstructure of Financial and Money Markets", CREST, 2006, and the 33rd Annual Meeting of the European Finance Association, EFA, 2006. This research was undertaken while Mancini visited the Department of Operations Research and Financial Engineering at Princeton University.

* Corresponding author at: Department of Economics, Princeton University, United States.
E-mail addresses: yacine@princeton.edu (Y. Aït-Sahalia), mancini@isb.uzh.ch (L. Mancini).

0304-4076/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2008.09.015

ε summarizes market microstructure effects, either informational or not: bid-ask bounces, discreteness of price changes, differences in trade sizes or informational content of price changes, gradual response of prices to a block trade, the strategic component of the order flow, inventory control effects, etc.

In the presence of market microstructure noise as in model (1.4), after suitable scaling, RV computed from the observed log-returns, $[Y,Y]_T$, is a consistent and asymptotically normal estimator of the quantity $2nE[\varepsilon^2]$. This quantity is not the object of interest, $\langle X,X\rangle_T$. In the high frequency limit where $\Delta \to 0$, market microstructure noise swamps the variance of the price signal.

Different solutions have been proposed for this problem. In the constant σ case, Zhou (1996) considers a bias correcting approach based on autocovariances that is similar to the estimator used in French et al. (1987). The behavior of this estimator has been studied by Zumbach et al. (2002). Efficient likelihood estimation of σ is studied by Aït-Sahalia et al. (2005), showing that incorporating ε explicitly in the likelihood function of the observed log-returns Y provides consistent, asymptotically normal and efficient estimators of the parameters that are robust to various forms of misspecifications of the noise term. Hansen and Lunde (2006) study the Zhou estimator and extensions in the case where volatility is time varying but conditionally nonrandom. Related contributions have been made by Oomen (2006) and Bandi and Russell (2006). The Zhou estimator and its extensions, however, are inconsistent. This means in this particular case that, as the frequency of observation increases, the estimator diverges instead of converging to $\langle X,X\rangle_T$: this was in fact recognized by Zhou, see p. 47 in Zhou (1996).

In the stochastic volatility case, Zhang et al. (2005b) propose a solution to this problem which makes use of the full data sample and delivers the first consistent estimators in the literature for $\langle X,X\rangle_T$, as well as $E[\varepsilon^2]$, in the presence of noise. The estimator, Two Scales Realized Volatility (TSRV), is based on subsampling, averaging and bias-correction.

Our objective in this paper is to examine and compare the out-of-sample performance of RV and TSRV as estimators of the quadratic variation, in a variety of contexts that are outside the controlled setup that has been used to derive the theoretical properties of these estimators. In other words, does the better performance of TSRV that is evident when looking at its theoretical properties extend to the out-of-sample forecasts when the data generating mechanism incorporates realistic features of the volatility process such as time series dependence in the noise, correlation of the noise with the price process, long memory, jumps and the like?

This paper complements other papers which have studied the forecasting ability of various volatility estimators. Andersen et al. (2003, 2004, 2005) study the standard RV estimator, looking at its pointwise forecasts and the efficiency loss relative to optimal, but unfeasible, forecasts based on the entire path of volatility. Their assumed model has an ARMA structure for the log of the volatility, and is one of the data generating processes we will consider. Under the eigenfunction stochastic volatility model of Meddahi (2001), they obtain analytical formulae for the autocovariance functions of RV. Market microstructure noise is not incorporated in those papers, however.

Analytical formulae for the autocovariance of RV with noise are given independently by Garcia and Meddahi (2006), and by Ghysels and Sinko (2006a) within the mixed data sampling (MIDAS) framework. The latter authors study the autocovariance corrections for RV proposed by Zhou (1996) and later extended by Hansen and Lunde (2006), and show that they do not improve volatility forecasting. Further results are reported in Ghysels and Sinko (2006b) and Ghysels et al. (forthcoming), whose data generating process for the volatility is of the same type as in the papers just cited, but explicitly include market microstructure noise. They find that at the high frequencies where market microstructure noise matters (although their study does not consider observation frequencies higher than one minute), TSRV predicts volatility the best; they also find that the noise component constructed as a difference between the standard RV estimator and TSRV, is predictable on average. While writing this paper, we became aware that Andersen et al. (2006) are in the process of extending their previous forecasting work to market microstructure robust noise measures.

Corradi et al. (forthcoming, 2005) construct conditional predictive densities and confidence intervals for IV and estimate these densities using nonparametric kernel estimators. They use different estimators for IV, including RV and TSRV and derive confidence bands around their kernel estimators. Their findings confirm the theoretical predictions of Zhang et al. (2005b) that, when the time interval between successive observations becomes small, the signal to noise ratio of the data decreases, and realized volatility and bipower variation tend to explode, instead of converging to the increments of quadratic variation. This result is visible from their predictive densities, as the ranges of the densities of these two uncorrected estimators widen considerably as the observation frequency increases. They also find that, TSRV, as a microstructure-robust measure of volatility, is stable, and increasing the frequency at which the data are sampled does not seem to induce any appreciable distortion in its density estimator.

Unlike those papers, we perform a horse race between different estimators by running Mincer–Zarnowitz forecasting regressions, and use widely different data generating processes. On the other hand, unlike Corradi et al. (forthcoming, 2005), but like the other papers mentioned, we look only at point estimates and not at the full forecast distribution.

Our results are unambiguous: no matter what data generating process we consider, and what assumptions we make on the noise (magnitude within reason, iid or autocorrelated, correlated or not with the price signal), we find that TSRV produces better forecasts of future IV than the standard RV does, sometimes by a wide margin.

This paper is organized as follows. Section 2 summarizes the RV and TSRV estimators we consider. Section 3 presents our Monte Carlo results under different assumptions: first the standard Heston stochastic volatility model; second a jump-diffusion model for the volatility; third a log-volatility model; fourth a heterogeneous autoregressive realized volatility (HAR-RV) model; fifth a long memory model for volatility based on a fractional Ornstein Uhlenbeck process. In Section 4, we study the forecasting ability of these estimators when applied to high frequency DJIA stocks data. Section 5 concludes.

2. Estimating the quadratic variation: RV and TSRV

We start with a brief description of the two estimators, RV and TSRV, and their theoretical properties. If one uses all the log-returns data available (say sampled every second),

$$[Y,Y]_T^{(\mathrm{all})} \overset{\mathcal{L}}{\approx} \underbrace{\langle X,X\rangle_T}_{\text{object of interest}} + \underbrace{2nE[\varepsilon^2]}_{\text{bias due to noise}} + \underbrace{\left[\underbrace{4nE[\varepsilon^4]}_{\text{due to noise}} + \underbrace{\frac{2T}{n}\int_0^T \sigma_t^4 \mathrm{d}t}_{\text{due to discretization}}\right]^{1/2}}_{\text{total variance}} Z_{\text{total}}, \tag{2.1}$$

conditionally on the X process, where $\overset{\mathcal{L}}{\approx}$ denotes stable convergence in law and Z denotes a standard normal variable. So the bias

term due to the noise, which is of order $O(n)$, swamps the true quadratic variation $\langle X, X\rangle_T$, which is of order $O(1)$.

Of course, sampling as prescribed by $[Y, Y]_T^{(\text{all})}$ is not what is done in practice. Instead, the estimator $[Y, Y]_T^{(\text{sparse})}$ constructed by summing squared log-returns at some lower frequency: 5 min, or 10, 15, 30 min, is typically used. Reducing the value of n, from say 23,400 (1 s sampling) to $n_{\text{sparse}} = 78$ ($\Delta_{\text{sparse}} = 5$ min sampling over the same 6.5 h), has the advantage of reducing the magnitude of the bias term $2nE[\varepsilon^2]$. Yet, one of the most basic lessons of statistics is that one should not do this.

TSRV is a simple method to tackle the problem: first, partition the original grid of observation times, $G = \{t_0, \ldots, t_n\}$ into subsamples, $G^{(k)}$, $k = 1, \ldots, K$, where $n/K \to \infty$ as $n \to \infty$. For example, for $G^{(1)}$ start at the first observation and take an observation every 5 min; for $G^{(2)}$, start at the second observation and take an observation every 5 min, etc. Then we average the estimators obtained on the subsamples. To the extent that there is a benefit to subsampling, this benefit can now be retained, while the variation of the estimator can be lessened by the averaging.

This gives rise to the estimator $[Y, Y]_T^{(\text{avg})} = \frac{1}{K}\sum_{k=1}^{K}[Y, Y]_T^{(k)}$ constructed by averaging the estimators $[Y, Y]_T^{(k)}$ obtained on K grids of average size $\bar{n} = n/K$. The properties of this estimator are given by

$$[Y, Y]_T^{(\text{avg})} \overset{\mathcal{L}}{\approx} \underbrace{\langle X, X\rangle_T}_{\text{object of interest}} + \underbrace{2\bar{n}E[\varepsilon^2]}_{\text{bias due to noise}} + \underbrace{\left[\underbrace{4\frac{\bar{n}}{K}E[\varepsilon^4]}_{\text{due to noise}} + \underbrace{\frac{4T}{3\bar{n}}\int_0^T \sigma_t^4 \mathrm{d}t}_{\text{due to discretization}}\right]^{1/2}}_{\text{total variance}} Z_{\text{total}}.$$

While a better estimator than $[Y, Y]_T^{(\text{all})}$, $[Y, Y]_T^{(\text{avg})}$ remains biased. The bias of $[Y, Y]_T^{(\text{avg})}$ is $2\bar{n}E[\varepsilon^2]$; of course, $\bar{n} < n$, so progress is being made. But one can go one step further. Indeed, $E[\varepsilon^2]$ can be consistently approximated using RV computed with all the observations:

$$\widehat{E[\varepsilon^2]} = \frac{1}{2n}[Y, Y]_T^{(\text{all})}. \tag{2.2}$$

Hence the bias of $[Y, Y]^{(\text{avg})}$ can be consistently estimated by $\frac{\bar{n}}{n}[Y, Y]_T^{(\text{all})}$. TSRV is the bias-adjusted estimator for $\langle X, X\rangle$ constructed as

$$\widehat{\langle X, X\rangle}_T^{(\text{tsrv})} = \underbrace{[Y, Y]_T^{(\text{avg})}}_{\text{slow time scale}} - \frac{\bar{n}}{n}\underbrace{[Y, Y]_T^{(\text{all})}}_{\text{fast time scale}}. \tag{2.3}$$

If the number of subsamples is optimally selected as $K^* = cn^{2/3}$, then TSRV has the following distribution:

$$\widehat{\langle X, X\rangle}_T^{(\text{tsrv})} \overset{\mathcal{L}}{\approx} \underbrace{\langle X, X\rangle_T}_{\text{object of interest}} + \frac{1}{n^{1/6}}\underbrace{\left[\underbrace{\frac{8}{c^2}E[\varepsilon^2]^2}_{\text{due to noise}} + \underbrace{c\frac{4T}{3}\int_0^T \sigma_t^4 \mathrm{d}t}_{\text{due to discretization}}\right]^{1/2}}_{\text{total variance}} Z_{\text{total}}. \tag{2.4}$$

Unlike all the previously considered ones, this estimator is now correctly centered. The constant c can be set to minimize the total asymptotic variance above.

In small samples, a small sample refinement to $\widehat{\langle X, X\rangle}_T$ can be constructed as follows

$$\widehat{\langle X, X\rangle}_T^{(\text{tsrv,adj})} = \left(1 - \frac{\bar{n}}{n}\right)^{-1}\widehat{\langle X, X\rangle}_T^{(\text{tsrv})}. \tag{2.5}$$

The difference with the estimator (2.3) is of order $O_p(K^{-1})$, and thus the two estimators have the same asymptotic behaviors to the order that we consider. However, the estimator (2.5) is unbiased to higher order.

When the microstructure noise ε is iid, the log-returns follow the MA(1) model

$$Y_{\tau_i} - Y_{\tau_{i-1}} = \int_{\tau_{i-1}}^{\tau_i}\mu_t \mathrm{d}t + \int_{\tau_{i-1}}^{\tau_i}\sigma_t \mathrm{d}W_t + \varepsilon_{\tau_i} - \varepsilon_{\tau_{i-1}}. \tag{2.6}$$

Hence the corresponding negative autocorrelation can affect the estimate of quadratic variations given by $[Y, Y]_T^{(\text{sparse})}$. An ad hoc remedy is to de-mean and filter the raw high frequency returns using an MA(1) model before applying such an estimator; see Andersen et al. (2001).

In the simulation and empirical work that follows, RV will refer to the estimator $[Y, Y]_T^{(\text{sparse})}$ (with or without the pre-MA(1) filtering of the raw data), at different values of the sparse sampling interval Δ_{sparse} and TSRV to the estimator $\widehat{\langle X, X\rangle}_T^{(\text{tsrv,adj})}$ (with no need for MA(1) filtering), with different values of the parameter K controlling the number of subgrids we use.

Finally, while TSRV provides the first consistent and asymptotic (mixed) normal estimator of the quadratic variation $\langle X, X\rangle_T$, as can be seen from (2.4), it has the rate of convergence $n^{-1/6}$. Zhang (2006) shows that it is possible to generalize TSRV to multiple time scales, by averaging not just on two time scales but on multiple time scales. For suitably selected weights, the resulting estimator, MSRV, converges to $\langle X, X\rangle_T$ at the slightly faster rate $n^{-1/4}$. Computing this estimator, however, requires additional computations due to the extra averaging step (over an asymptotically increasing number of time scales) and will in general produce close results to those given by TSRV. An alternative derivation of TSRV and MSRV, based on kernels, has subsequently been proposed by Barndorff-Nielsen et al. (2006).

3. Monte Carlo evidence using alternative data generating processes

In this section we perform several experiments using widely different volatility models. In each experiment we compare the in- and out-of-sample forecast performances of RV and TSRV computed at different frequencies.

3.1. Heston stochastic volatility model

In this Monte Carlo experiment, we use as the data generating process the stochastic volatility model of Heston (1993) for the instantaneous variance

$$\begin{aligned}\mathrm{d}X_t &= (\mu - \sigma_t^2/2)\mathrm{d}t + \sigma_t \mathrm{d}W_{1,t}\\ \mathrm{d}\sigma_t^2 &= \kappa(\alpha - \sigma_t^2)\mathrm{d}t + \gamma\sigma_t \mathrm{d}W_{2,t}.\end{aligned} \tag{3.1}$$

We set the parameters $\mu, \kappa, \alpha, \gamma$ and ρ, the correlation coefficient between the two Brownian motions W_1 and W_2, to parameter values which are reasonable for a stock price, as in Zhang et al. (2005b), namely, $\mu = 0.05$, $\kappa = 5$, $\alpha = 0.04$, $\gamma = 0.5$, $\rho = -0.5$. The volatility parameters satisfy the Feller's condition $2\kappa\alpha \geq \gamma^2$ which makes the zero boundary unattainable by the volatility process. We simulate $M = 10{,}000$ sample paths of the process using the Euler scheme at a time interval $\Delta = 1$ s. We

Table 1
In-sample estimates of daily $IV \times 10^4$ based on 10,000 simulated paths under the Heston model, $d\sigma^2 = 5(0.04 - \sigma^2)dt + 0.5\,\sigma\,dW_2$

	Bias	Var	RMSE	Rel. bias	Rel. var	Rel. RMSE
RV 5 min	1.560	0.318	1.659	2.508	47.701	7.348
RV 5 min MA(1)	1.348	0.274	1.446	2.026	26.985	5.576
TSRV 5 min	−0.014	0.071	0.266	−0.011	0.022	0.149
RV 10 min	0.779	0.390	0.999	1.238	10.664	3.492
TSRV 10 min	−0.032	0.135	0.369	−0.023	0.035	0.188
RV 15 min	0.528	0.474	0.867	0.834	5.376	2.464
TSRV 15 min	−0.050	0.199	0.449	−0.034	0.051	0.227
RV 30 min	0.275	0.780	0.925	0.422	1.522	1.304
TSRV 30 min	−0.110	0.395	0.638	−0.072	0.097	0.319
TSRV minimum variance	−0.001	0.020	0.140	−0.004	0.010	0.099

The efficient log-price $dX = (0.05 - \sigma^2/2)dt + \sigma\,dW_1$, and correlation $\rho = -0.5$ between Brownian motions. The observed log-price $Y = X + \varepsilon$, where $\varepsilon \sim NID(0, 0.001^2)$. Euler discretization scheme with time step $\Delta = 1$ s.

assume that the market microstructure noise, ε, has a Gaussian distribution, and $(E\varepsilon^2)^{1/2} = 0.001$, i.e. the standard deviation of the noise is 0.1% of the value of the asset price.

We simulate "continuous" sample paths of $m + 1 = 101$ days, $[0, T_1], \ldots, [T_{99}, T_{100}], [T_{100}, T_{101}]$, for the observed log-price process Y, that is each sample path consists of $101 \times 23{,}400$ log-returns, assuming that each trading day consists of 6.5 h, as is the case on the NYSE and NASDAQ.

When constructing the RV estimator, the first $m = 100$ days are used to estimate an MA(1) filter in order to remove as much of the noise as possible, and consistently with the common practice in the empirical RV literature. This step is not needed for the TSRV estimator. On each simulated sample path, the MA(1) model is estimated using $100 \times 78 = 7800$ log-returns at the 5 min frequency and the de-meaned MA(1) filtered returns on the 100th day are used to estimate $\int_{T_{m-1}}^{T_m} \sigma_t^2 dt$; a similar procedure is applied, for instance, by Andersen et al. (2001) to investigate the IV's of the DJIA stocks. The 100th day is used for the in-sample estimate of the IV and the last 101st day is saved for the out-of-sample forecast of the IV.

3.1.1. One day ahead forecasts of IV

Denoting by $T_1 - T_0 = T_2 - T_1$ the one day time interval, it follows that

$$E[\sigma_{T_1}^2 | \mathcal{F}_{T_0}] = a\,\sigma_{T_0}^2 + b, \tag{3.2}$$

where $a = e^{-\kappa(T_1 - T_0)}$, $b = \alpha(1 - e^{-\kappa(T_1 - T_0)})$, and σ_t^2 follows the model (3.1): see e.g., Cox et al. (1985). $\mathcal{F}_T = \sigma\{\sigma_t^2; t \le T\}$ is the information set or σ-field generated by the instantaneous variance process up to time T. Interchanging the integration operators, we have

$$E\left[\int_{T_0}^{T_1} \sigma_t^2 dt | \mathcal{F}_{T_0}\right] = a_{\text{int}}\sigma_{T_0}^2 + b_{\text{int}}, \tag{3.3}$$

where $a_{\text{int}} = 1/\kappa(1 - e^{-\kappa(T_1 - T_0)})$ and $b_{\text{int}} = \alpha(T_1 - T_0) - \alpha/\kappa(1 - e^{-\kappa(T_1 - T_0)})$. Then using (3.3) and (3.2) gives

$$\begin{aligned}
&E\left[E\left[\int_{T_1}^{T_2} \sigma_t^2 dt | \mathcal{F}_{T_1}\right] | \mathcal{F}_{T_0}\right] \\
&= E[a_{\text{int}}\sigma_{T_1}^2 + b_{\text{int}} | \mathcal{F}_{T_0}] = a_{\text{int}} E[\sigma_{T_1}^2 | \mathcal{F}_{T_0}] + b_{\text{int}} \\
&= a_{\text{int}}(a\,\sigma_{T_0}^2 + b) + b_{\text{int}} = a(a_{\text{int}}\sigma_{T_0}^2) + a_{\text{int}}b + b_{\text{int}} \\
&= a\left(E\left[\int_{T_0}^{T_1} \sigma_t^2 dt | \mathcal{F}_{T_0}\right] - b_{\text{int}}\right) + a_{\text{int}}b + b_{\text{int}} \\
&E\left[\int_{T_1}^{T_2} \sigma_t^2 dt | \mathcal{F}_{T_0}\right] = aE\left[\int_{T_0}^{T_1} \sigma_t^2 dt | \mathcal{F}_{T_0}\right] + b(T_1 - T_0),
\end{aligned} \tag{3.4}$$

which will give the out-of-sample forecast Eq. (3.6).

Using (3.4), Bollerslev and Zhou (2002) propose a GMM-type estimator for the Heston model based on intra-day returns for the foreign exchange markets. Notice that the exact conditional forecast of $\int_{T_1}^{T_2} \sigma_t^2 dt$ is

$$E\left[\int_{T_1}^{T_2} \sigma_t^2 dt | \mathcal{F}_{T_1}\right] = a_{\text{int}}\sigma_{T_1}^2 + b_{\text{int}}, \tag{3.5}$$

which is not feasible as $\sigma_{T_1}^2$ is not observed.

3.1.2. In-sample estimations of IV

We present the in-sample estimation of IV given by TSRV and RV under the Heston volatility model. The same procedure is adopted in the other experiments. We estimate the IV on the 100th day, $\int_{T_{m-1}}^{T_m} \sigma_t^2 dt$, using the RV estimator $[Y, Y]^{(\text{sparse})}$ and the TSRV estimator $\widehat{\langle X, X\rangle}^{(\text{tsrv,adj})}$ based on unfiltered returns and arbitrarily-determined frequencies of 5, 10, 15 and 30 min. We also report the results for $[Y, Y]^{(\text{sparse})}$ based on de-meaned MA(1) filtered 5 min returns and $\widehat{\langle X, X\rangle}^{(\text{tsrv,adj})}$ under the optimal sampling frequency.

Table 1 shows the in-sample results for the different estimation strategies and TSRV estimators always largely outperform all RV estimators in terms of bias, variance and RMSE at each frequency. In-sample simulations for this model are also reported in Zhang et al. (2005b), but with half variance for the microstructure noise, that is $(E\varepsilon^2)^{1/2} = 0.0005$. Hence the outperformance of TSRV over RV is robust to an increase in the magnitude of the noise, as expected. Of course, going in the reverse direction by decreasing the amount of noise diminishes the advantage of TSRV over RV, but empirical studies such as those cited in Aït-Sahalia et al. (2005) suggest level of noise that are significantly higher than 0.0005. Table 1 also shows that the MA(1) filter tends to improve the estimation performance of RV estimators, which remains largely inferior to the TSRV performance. Similar in-sample results are obtained for all the subsequent Monte Carlo experiments and to save space they are not reported here, but collected in Aït-Sahalia and Mancini (2007).

The minimum variance TSRV is computed using $K \approx 100$ subsamples which correspond to a slow time scale of less than two minutes.[1] Fig. 1 shows the finite sample distribution of the minimum variance TSRV estimator (2.5) and the corresponding asymptotic Gaussian distribution. Incidentally, the two distributions appear to be very close, confirming the finding of Zhang et al. (2005a), that are robust to an increase in the magnitude of microstructure noise. In all the subsequent Monte Carlo experiments the finite sample distributions of the minimum variance TSRV are very close to the asymptotic distribution and the corresponding plots will be omitted.

[1] Roughly the same number of subsamples is also used in the subsequent Monte Carlo experiments.

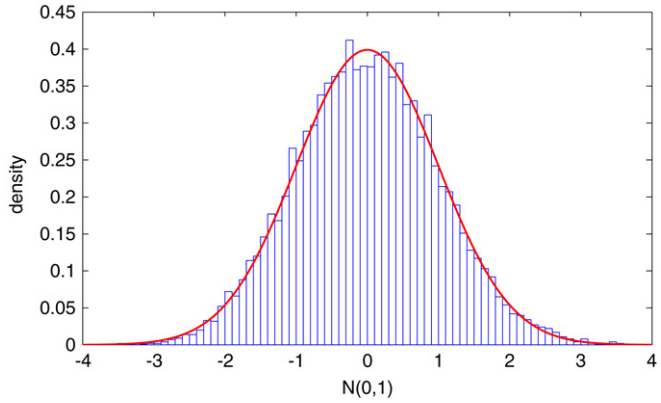

Fig. 1. Asymptotic and small sample distributions of the minimum variance TSRV estimator based on 10,000 simulated paths of the instantaneous variance under the Heston model, $d\sigma^2 = 5(0.04 - \sigma^2)dt + 0.5\sigma dW_2$. The efficient log-price $dX = (0.05 - \sigma^2/2)dt + \sigma dW_1$, and correlation $\rho = -0.5$ between Brownian motions. The observed log-price $Y = X + \varepsilon$, where $\varepsilon \sim NID(0, 0.001^2)$. Euler discretization scheme with time step $\Delta = 1$ s.

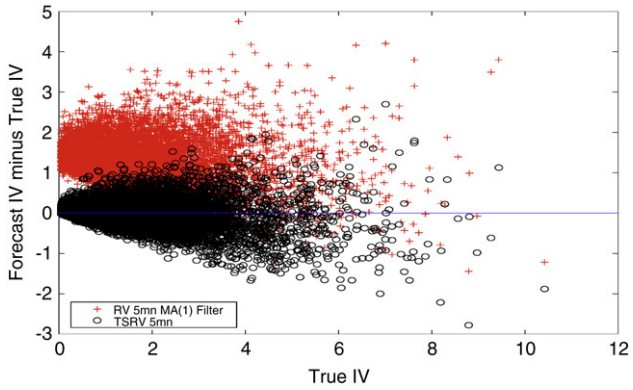

Fig. 2. Out-of-sample forecast error, (i.e. forecast minus true value of $IV \times 10^4$), plotted versus true value of $IV \times 10^4$ for the 10,000 simulated paths of the instantaneous variance under the Heston model, $d\sigma^2 = (0.04-\sigma^2)dt+0.5\sigma dW_2$. The efficient log-price $dX = (0.05 - \sigma^2/2)dt + \sigma dW_1$, and correlation $\rho = -0.5$ between Brownian motions. The observed log-price $Y = X + \varepsilon$, where $\varepsilon \sim NID(0, 0.001^2)$. Euler discretization scheme with time step $\Delta = 1$ s.

3.1.3. Out-of-sample forecasts of IV

In this section we compare the one day ahead out-of-sample forecasts of the IV given by TSRV and RV. As discussed in Section 3.1.1, the relation between the conditional means of IV is

$$E\left[\int_{T_m}^{T_{m+1}} \sigma_t^2 dt | \mathcal{F}_{T_{m-1}}\right] = e^{-\kappa D} E\left[\int_{T_{m-1}}^{T_m} \sigma_t^2 dt | \mathcal{F}_{T_{m-1}}\right] + \alpha(1 - e^{-\kappa D})D, \tag{3.6}$$

where $D = T_{m+1} - T_m = T_m - T_{m-1}$ is one day on an annualized base. Replacing the conditional mean in the right-hand-side by the estimate of the IV on day m provides a simple method to forecast the IV on day $m+1$. Precisely we proceed as follows. We split each simulated sample path in two parts. The first part of $100 \times 23{,}400$ high frequency log-returns is used to estimate the time series of 100 daily IV and an AR(1) model for the IV itself or equivalently to estimate intercept and slope in the forecast Eq. (3.6) where the conditional expectation in the right-hand-side is replaced by the estimated IV. The second part of the simulated sample path consists of 23,400 high frequency log-returns, it represents the last 101st trading day of high frequency data and it is saved for out-of-sample purposes. The out-of-sample forecast of IV on the day 101st is provided by the AR(1) model for IV and compared with true IV realized on that day. This procedure is repeated for each sample path and for both TSRV and RV computed at the different frequencies. In empirical applications the true underlying model parameters are unknown and finite sample properties of stochastic volatility models, such as mean reversion, can be quite different from the corresponding asymptotic properties. Hence estimation of Eq. (3.6) is required to be realistic.

In the tradition of Mincer and Zarnowitz (1969) and Chong and Hendry (1986), we compare the alternative volatility forecasts by projecting the true realized IV on day $m+1$, $\int_{T_m}^{T_{m+1}} \sigma_t^2 dt$, on a constant and the various model forecasts. The forecast evaluation regressions take the form

$$\left\{\int_{T_m}^{T_{m+1}} \sigma_t^2 dt\right\}_j = b_0 + b_1\{\mathrm{TSRV}_{T_{m+1}|T_m}\}_j + b_2\{\mathrm{RV}_{T_{m+1}|T_m}\}_j + \mathrm{error}_j, \tag{3.7}$$

for $j = 1, \ldots, 10{,}000$. $\mathrm{TSRV}_{T_{m+1}|T_m}$ is the out-of-sample forecast of IV from day m to day $m+1$ given by the AR(1) model for IV estimated using the time series of TSRV daily IV. $\mathrm{RV}_{T_{m+1}|T_m}$ is similarly computed.

The OLS estimates of the forecast evaluation regressions (3.7) are reported in Table 2. In all cases TSRV forecasts largely outperform RV forecasts; Fig. 2 visually confirms this result. Moreover, in almost all cases, adding any RV forecasts to TSRV forecasts gives no additional explanatory power, that is the R^2 of the extended regressions do not increase. The only exception is the extreme situation of TSRV computed at the (excessively long) 30 min frequency. In all cases, the null hypothesis $H_0 : b_0 = 0$ and $H_0 : b_1 = 1$ are rejected. However, also using the true in-sample IV on day $m = 100$ to forecast the IV on day $m+1$, the previous hypothesis are rejected. This is mainly due to a Jensen's inequality effect which links in a nonlinear way the parameters of the Heston model and the IV process.

We performed a number of robustness checks to support the previous results. To save space, and because the results are consistent with those above, these additional results are not reported in the present paper, but collected in the supplementary appendix (Aït-Sahalia and Mancini, 2007). For instance, we investigate the out-of-sample forecasts of the integrated volatility $IV^{1/2}$ instead of the integrated variance, IV. Then in the forecast evaluation regressions (3.7), IV and the corresponding estimates are replaced by $IV^{1/2}$. The overall conclusions are the same as in Table 2, meaning that the previous findings are robust to Jensen's inequality effects. We also repeat the previous simulation study using a nearly integrated Heston model to approximate long memory features exhibited in the real world by the volatility of some assets; see Section 3.5 below for a more theoretically-based approach to long memory. In the Heston model, we set the mean reversion coefficient $\kappa = 1$ and the local volatility parameter $\gamma = 0.275$ in order to satisfy Feller's condition. The findings of the robustness checks confirm the results reported above. The main difference is that RV has an even larger relative variance when estimating IV in-sample. This is due to a larger over-estimation of small integrated variances.

3.2. Jump-diffusion model

Empirically it has been observed that the Heston model has some difficulties in fitting asset and option prices (see for instance Jones (2003)) and hence it has been extended in several directions, in particular by including jump components to the volatility process. To investigate the impact of jumps on RV and TSRV, we now consider a model where the log-price dynamic

Table 2
Out-of-sample, one day ahead forecasts of daily IV in percentage based on 10,000 simulated paths under the Heston model, $d\sigma^2 = 5\,(0.04 - \sigma^2)dt + 0.5\,\sigma\,dW_2$

	b_0	b_1	b_2	R^2
RV 5 min	−1.467 (0.016)		0.971 (0.005)	0.809
RV 5 min MA(1)	−1.105 (0.013)		0.916 (0.004)	0.841
RV 10 min	−0.689 (0.013)		0.958 (0.005)	0.782
RV 15 min	−0.432 (0.013)		0.950 (0.005)	0.753
RV 30 min	−0.136 (0.014)		0.921 (0.006)	0.670
TSRV 5 min	0.018 (0.006)	0.991 (0.003)		0.928
TSRV 5 min + RV 5 min	0.146 (0.016)	1.057 (0.008)	−0.074 (0.009)	0.928
TSRV 5 min + RV 5 min MA(1)	0.190 (0.014)	1.116 (0.010)	−0.126 (0.010)	0.929
TSRV 5 min + RV 10 min	0.159 (0.010)	1.120 (0.008)	−0.146 (0.008)	0.930
TSRV 5 min + RV 15 min	0.119 (0.008)	1.107 (0.007)	−0.134 (0.007)	0.930
TSRV 5 min + RV 30 min	0.089 (0.007)	1.085 (0.006)	−0.118 (0.006)	0.930
TSRV 10 min	0.023 (0.007)	0.996 (0.004)		0.888
TSRV 10 min + RV 5 min	−0.184 (0.019)	0.884 (0.010)	0.122 (0.011)	0.890
TSRV 10 min + RV 5 min MA(1)	−0.115 (0.018)	0.891 (0.013)	0.103 (0.013)	0.889
TSRV 10 min + RV 10 min	0.114 (0.013)	1.087 (0.011)	−0.098 (0.011)	0.889
TSRV 10 min + RV 15 min	0.148 (0.010)	1.158 (0.010)	−0.178 (0.011)	0.891
TSRV 10 min + RV 30 min	0.122 (0.008)	1.151 (0.008)	−0.183 (0.008)	0.893
TSRV 15 min	0.026 (0.008)	1.005 (0.004)		0.853
TSRV 15 min + RV 5 min	−0.507 (0.020)	0.708 (0.011)	0.316 (0.011)	0.864
TSRV 15 min + RV 5 min MA(1)	−0.482 (0.019)	0.601 (0.014)	0.386 (0.013)	0.864
TSRV 15 min + RV 10 min	−0.065 (0.014)	0.909 (0.013)	0.101 (0.013)	0.854
TSRV 15 min + RV 15 min	0.085 (0.012)	1.091 (0.013)	−0.090 (0.013)	0.853
TSRV 15 min + RV 30 min	0.138 (0.010)	1.211 (0.010)	−0.230 (0.011)	0.859
TSRV 30 min	0.071 (0.011)	1.014 (0.006)		0.759
TSRV 30 min + RV 5 min MA(1)	−0.971 (0.017)	0.158 (0.013)	0.790 (0.011)	0.843
TSRV 30 min + RV 30 min	0.085 (0.013)	1.047 (0.017)	−0.033 (0.017)	0.759
TSRV minimum variance	0.005 (0.004)	0.993 (0.002)		0.961

The efficient log-price $dX = (0.05 - \sigma^2/2)dt + \sigma\,dW_1$, and correlation $\rho = -0.5$ between Brownian motions. The observed log-price $Y = X + \varepsilon$, where $\varepsilon \sim NID(0, 0.001^2)$. Euler discretization scheme with time step $\Delta = 1$ s. OLS standard errors in parenthesis. Eq. (3.6) is estimated using the previous 100 simulated days and regressing the corresponding estimates of $\int_{T_i}^{T_{i+1}} \sigma_t^2 dt$ on a constant and $\int_{T_{i-1}}^{T_i} \sigma_t^2\,dt$, for $i = 1, \ldots, 99$.

is given by Eq. (3.1), but the volatility follows a jump-diffusion Heston type model

$$d\sigma_t^2 = \kappa(\alpha - \sigma_{t_-}^2)dt + \gamma\sigma_{t_-}\,dW_{2,t} + J_t\,dq_t, \tag{3.8}$$

where $\sigma_{t_-}^2 = \lim_{s\uparrow t}\sigma_s^2$, q_t is a Poisson process with intensity λ and J_t is the jump size, assumed to be exponentially distributed with parameter ξ. The processes q and J are independent of the two Brownian motions. Without jumps, $\lambda = 0$, Eq. (3.8) reduces to the Heston's model (3.1). The long run mean of the volatility process is $E[\sigma_t^2] = \alpha + \xi\lambda/\kappa$.[2]

We set the jump coefficients $\lambda = \Delta \times 23{,}400/2$ which implies (on average) two volatility jumps per day, and $\xi = 0.0007$, that is the jump size is about 2% of the unconditional variance. The diffusion coefficients are as in Section 3.1, but $\alpha = 0.035$.

In-sample estimation results are rather close to the findings under the Heston model and reported in Aït-Sahalia and Mancini (2007). Hence, as in the previous experiment, TSRV outperforms RV in terms of bias, variance and RMSE at each frequency. Table 3 shows the out-of-sample simulation results. Again as in the previous cases, TSRV provides more efficient and unbiased forecasts than RV and at each frequency; TSRV results in larger R^2 than RV.

It is known that empirically asset returns tend to display a jump component. Hence we simulate high frequency data using the same Heston volatility model as in Section 3.1, but adding a jump component to the log-price process X. All simulation results are collected in Aït-Sahalia and Mancini (2007) and show that TSRV outperforms RV also in that setting.

[2] This can be seen integrating Eq. (3.8) and taking the unconditional expectation

$$E[\sigma_t^2] = E[\sigma_0] + E\left[\int_0^t \kappa(\alpha - \sigma_{s_-}^2)ds\right] + E\left[\int_0^t \gamma\sigma_{s_-}\,dW_{2,s}\right] + E\left[\sum_{j=1}^{q_t} J_j\right].$$

The Itô integral has zero mean, and $E[\sum_{j=1}^{q_t} J_j] = \xi\lambda t$. Denoting $E[\sigma_t^2] = \bar{\sigma}^2$ and using Fubini's theorem gives $\bar{\sigma}^2 = \bar{\sigma}^2 + \kappa(\alpha - \bar{\sigma}^2)\,t + \xi\lambda t$, which implies $\bar{\sigma}^2 = \alpha + \xi\lambda/\kappa$.

3.3. Log-volatility model

In empirical work, volatility is sometimes modeled and predicted with ARMA models that are constructed using logs of realized volatility; see Andersen et al. (2001, 2003) for the use of log-RV for stock return and exchange rate volatility. The rationale for doing this is that the distribution of log-RV can be closer to Gaussian than that of RV. Goncalves and Meddahi (2005b) generalize this to the use of Box–Cox transforms. Without log or other transformations, Goncalves and Meddahi (2005a) and Zhang et al. (2005a) provide Edgeworth expansions for RV and TSRV with and without noise.

Here, it turns out that little of the difference in predictability can be attributed to non-normality. We run simulations using a log-volatility model calibrated to the parameter values reported by these papers and add the same tiny amount of microstructure noise as above. Let $IV_t^{(d)}$ denote the daily integrated variance $IV_t^{(d)} = \int_{t-d}^t \sigma_s^2 ds$, where $d = 1$ day. Following Andersen et al. (2003), the daily logarithmic standard deviation l_t is assumed to evolve according to an AR(5) model[3]

$$l_t = \frac{1}{2}\log(IV_t^{(d)}) = \phi_0 + \sum_{i=1}^{5}\phi_i l_{t-id} + u_t, \tag{3.9}$$

where u_t is a strong white noise, $u_t \sim NID(0, Eu^2)$. In contrast to the previous volatility models, the log-volatility model (3.9) does not specify the instantaneous variance dynamic and is defined in discrete time. Intraday efficient log-returns r_t are given by the daily integrated volatility times a noise term

$$r_t = \sqrt{IV_t^{(d)}}\,z_t, \tag{3.10}$$

[3] Precisely, to capture long memory features in IV, Andersen et al. (2003) use a fractionally integrated AR process. We investigate the long memory issue in Sections 3.4 and 3.5.

Table 3
Out-of-sample forecasts, one day ahead of daily IV in percentage based on 10,000 simulated paths under the Heston jump-diffusion model, $d\sigma^2 = 5(0.035 - \sigma^2)dt + 0.5\sigma dW_2 + J\,dq$, q is a Poisson process with intensity $\lambda = \Delta \times 23{,}400/2$ and $J \sim \text{Exp}(0.0007)$

	b_0	b_1	b_2	R^2
RV 5 min	−1.298 (0.020)		0.962 (0.005)	0.806
RV 5 min MA(1)	−1.026 (0.018)		0.926 (0.004)	0.827
RV 10 min	−0.481 (0.019)		0.945 (0.005)	0.763
RV 15 min	−0.197 (0.019)		0.934 (0.006)	0.725
RV 30 min	0.170 (0.021)		0.903 (0.007)	0.630
TSRV 5 min	0.093 (0.009)	0.983 (0.003)		0.910
TSRV 5 min + RV 5 min	0.160 (0.019)	1.017 (0.009)	−0.038 (0.010)	0.911
TSRV 5 min + RV 5 min MA(1)	0.233 (0.018)	1.077 (0.011)	−0.097 (0.011)	0.911
TSRV 5 min + RV 10 min	0.226 (0.013)	1.100 (0.008)	−0.133 (0.009)	0.912
TSRV 5 min + RV 15 min	0.188 (0.011)	1.082 (0.007)	−0.117 (0.008)	0.912
TSRV 5 min + RV 30 min	0.160 (0.010)	1.054 (0.006)	−0.092 (0.006)	0.912
TSRV 10 min	0.145 (0.012)	0.983 (0.004)		0.860
TSRV 10 min + RV 5 min	−0.283 (0.023)	0.754 (0.011)	0.247 (0.011)	0.866
TSRV 10 min + RV 5 min MA(1)	−0.196 (0.022)	0.743 (0.014)	0.240 (0.014)	0.864
TSRV 10 min + RV 10 min	0.180 (0.017)	1.017 (0.012)	−0.037 (0.012)	0.860
TSRV 10 min + RV 15 min	0.246 (0.014)	1.107 (0.011)	−0.137 (0.011)	0.862
TSRV 10 min + RV 30 min	0.233 (0.013)	1.097 (0.008)	−0.138 (0.009)	0.863
TSRV 15 min	0.184 (0.013)	0.988 (0.005)		0.817
TSRV 15 min + RV 5 min	−0.619 (0.023)	0.554 (0.011)	0.460 (0.011)	0.843
TSRV 15 min + RV 5 min MA(1)	−0.573 (0.022)	0.444 (0.014)	0.533 (0.013)	0.843
TSRV 15 min + RV 10 min	−0.019 (0.018)	0.782 (0.014)	0.217 (0.014)	0.821
TSRV 15 min + RV 15 min	0.188 (0.017)	0.993 (0.014)	−0.005 (0.014)	0.817
TSRV 15 min + RV 30 min	0.276 (0.015)	1.130 (0.011)	−0.163 (0.011)	0.820
TSRV 30 min	0.332 (0.017)	0.987 (0.006)		0.710
TSRV 30 min + RV 5 min MA(1)	−0.936 (0.020)	0.114 (0.012)	0.836 (0.010)	0.829
TSRV 30 min + RV 30 min	0.291 (0.019)	0.908 (0.017)	0.083 (0.017)	0.710
TSRV minimum variance	0.041 (0.006)	0.987 (0.002)		0.958

The efficient log-price $dX = (0.05 - \sigma^2/2)dt + \sigma dW_1$, and correlation $\rho = -0.5$ between Brownian motions. The observed log-price $Y = X + \varepsilon$, where $\varepsilon \sim NID(0, 0.001^2)$. Euler discretization scheme with time step $\Delta = 1$ s. OLS standard errors in parenthesis. The corresponding Eq. (3.6) is estimated using the previous 100 days of estimated IV, regressing the corresponding estimates of $\int_{T_i}^{T_{i+1}} \sigma_t^2 dt$ on a constant and $\int_{T_{i-1}}^{T_i} \sigma_t^2 dt$, for $i = 1, \ldots, 99$.

where $z_t \sim NID(0, 1)$. In our simulation experiment we set the parameter $\phi_0 = -0.0161$, $\phi_1 = 0.35$, $\phi_2 = 0.25$, $\phi_3 = 0.20$, $\phi_4 = 0.10$, $\phi_5 = 0.09$, and $(E\,u^2)^{1/2} = 0.02$ to induce approximately the same level of daily logarithmic standard deviations documented by Andersen et al. (2001) for the DJIA stocks. We simulate 10,000 sample paths of 201 × 23,400 log-returns, i.e. each sample path represents 201 days of log-returns sampled at one second intervals. The first 200 days are used to estimate the log-volatility model (3.9) and the last day is saved for the out-of-sample forecasts. In the current experiment we simulate longer sample paths (201 days instead of 101 days) than in the previous experiments to achieve a reasonable sample size for the estimation of the log-volatility model. The out-of-sample forecasts of IV are obtained by Eq. (3.9). As the logarithmic standard deviation l_t is updated on a daily base, AR-type forecast equations are correctly specified and directly implied by the assumed log-volatility model.

In-sample estimates of IV are reported in Aït-Sahalia and Mancini (2007) and show that TSRV outperforms RV in terms of bias, variance and RMSE as in the previous simulation experiments. Table 4 shows the out-of-sample forecasts of IV and largely confirms the previous findings. For instance, RV forecasts have no additional explanatory power in terms of R^2 when added to TSRV forecasts in the Mincer–Zarnowitz regressions.

3.4. Heterogeneous autoregressive realized volatility model

In this section, we simulate the integrated volatility process using the heterogeneous autoregressive realized volatility (HAR-RV) model proposed by Corsi (2004). The model is inspired by the heterogeneity of agents operating in financial markets, and in particular to the different time horizons (such as daily or monthly) relevant for different investors (such as day traders or fund managers); see also Müller et al. (1997) and Dacorogna et al. (1998). Such heterogeneity could explain the strong positive correlation between volatility and market participation observed empirically. Indeed, if all the investors were equal, the more agents acted in the market, the faster the price converged to its market value on which all agents agreed. Thus, the volatility had to be negatively correlated with market participation.

Another aspect captured by the HAR-RV model is that short term traders are concerned about long term volatility as it potentially affects expected future trends and riskiness. Hence on the one hand they react to changes in long term volatility by revising their trading strategies and causing volatility. On the other hand, the level of short term volatility does not affect the trading strategies of long term investors such as pension fund mangers or institutional investors. Such an asymmetric dependence among volatilities at different time scales can be formalized using a cascade volatility model from low frequencies to high frequencies. To simplify the model we will consider only three frequencies, i.e. daily, weekly, and monthly as in Corsi (2004). In the following, irrespective of the frequency, all the quantities are expressed on an annual basis. As in the previous section, $IV_t^{(d)}$ denotes the daily integrated variance and $RV_t^{(d)}$ the daily realized variance $RV_t^{(d)} = \sum_{j=0}^{d/\Delta} r_{t-j\Delta}^2$, where $r_{t-j\Delta} = x_{t-j\Delta} - x_{t-(j+1)\Delta}$, x_t is the true underlying log-price and $\Delta = 1$ s. Weekly and monthly realized variances, $RV^{(w)}$ and $RV^{(m)}$, are similarly defined by summing up squared log-returns over the last week and month, respectively. The HAR-RV model for the daily integrated variance is

$$IV_{t+\Delta}^{(d)} = \beta_0 + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + \omega_{t+\Delta}^{(d)}, \quad (3.11)$$

where $\omega_t^{(d)} \sim NID(0, \sigma_\omega^2)$. Similarly to the log-volatility model (3.9), the HAR-RV model does not specify the instantaneous variance dynamic. Hence intraday efficient log-returns, r_t, are given by Eq. (3.10).

The Eq. (3.11) has a simple AR structure and formally $IV^{(d)}$ does not belong to the class of long memory processes. However, the integrated variance can exhibit an apparent long memory behavior with an autocorrelogram decaying (approximately) hyperbolically

Table 4
Out-of-sample, one day ahead forecasts of daily IV in percentage based on 10,000 simulated paths under the log-volatility model (3.9), $l_t = -0.0161 + 0.35\, l_{t-d} + 0.25\, l_{t-2d} + 0.20\, l_{t-3d} + 0.10\, l_{t-4d} + 0.09\, l_{t-5d} + u_t$, where $u \sim NID(0, 0.02^2)$ and $d = 1$ day

	b_0	b_1	b_2	R^2
RV 5 min	−1.363 (0.025)		0.953 (0.008)	0.580
RV 5 min MA(1)	−1.142 (0.021)		0.941 (0.007)	0.645
RV 10 min	−0.592 (0.020)		0.950 (0.009)	0.553
RV 15 min	−0.280 (0.018)		0.927 (0.009)	0.514
RV 30 min	0.171 (0.017)		0.838 (0.010)	0.415
TSRV 5 min	−0.038 (0.008)	1.050 (0.005)		0.828
TSRV 5 min + RV 5 min	−0.084 (0.019)	1.032 (0.009)	0.024 (0.009)	0.828
TSRV 5 min + RV 5 min MA(1)	−0.040 (0.018)	1.049 (0.010)	0.001 (0.010)	0.828
TSRV 5 min + RV 10 min	0.013 (0.013)	1.084 (0.009)	−0.045 (0.009)	0.829
TSRV 5 min + RV 15 min	0.008 (0.011)	1.087 (0.008)	−0.051 (0.009)	0.829
TSRV 5 min + RV 30 min	−0.013 (0.009)	1.073 (0.007)	−0.035 (0.008)	0.829
TSRV 10 min	−0.037 (0.010)	1.073 (0.006)		0.752
TSRV 10 min + RV 5 min	−0.367 (0.022)	0.929 (0.011)	0.177 (0.011)	0.758
TSRV 10 min + RV 5 min MA(1)	−0.319 (0.021)	0.897 (0.013)	0.189 (0.012)	0.758
TSRV 10 min + RV 10 min	−0.084 (0.016)	1.037 (0.012)	0.044 (0.012)	0.752
TSRV 10 min + RV 15 min	−0.027 (0.013)	1.083 (0.011)	−0.012 (0.012)	0.752
TSRV 10 min + RV 30 min	−0.021 (0.011)	1.092 (0.009)	−0.025 (0.010)	0.752
TSRV 15 min	−0.017 (0.011)	1.085 (0.007)		0.694
TSRV 15 min + RV 5 min	−0.589 (0.024)	0.823 (0.012)	0.309 (0.012)	0.715
TSRV 15 min + RV 5 min MA(1)	−0.562 (0.022)	0.724 (0.014)	0.371 (0.013)	0.718
TSRV 15 min + RV 10 min	−0.211 (0.017)	0.925 (0.013)	0.187 (0.013)	0.700
TSRV 15 min + RV 15 min	−0.077 (0.015)	1.020 (0.013)	0.077 (0.013)	0.695
TSRV 15 min + RV 30 min	−0.013 (0.013)	1.092 (0.011)	−0.008 (0.011)	0.694
TSRV 30 min	0.103 (0.013)	1.078 (0.009)		0.571
TSRV 30 min + RV 5 min MA(1)	−0.900 (0.021)	0.419 (0.014)	0.657 (0.012)	0.673
TSRV 30 min + RV 30 min	0.034 (0.015)	0.953 (0.016)	0.142 (0.014)	0.575
TSRV minimum variance	−0.016 (0.006)	1.019 (0.003)		0.895

The efficient log-return $r_t = X_t - X_{t-\Delta} = \sqrt{IV_t^{(d)}} z_t$, where $z \sim NID(0, 1)$ and $\Delta = 1$ s. The observed log-price $Y = X + \varepsilon$, where $\varepsilon \sim NID(0, 0.001^2)$. OLS standard errors in parenthesis. Eq. (3.9) is estimated using the previous 200 days of estimated daily IV.

as shown for instance in Fig. 3[4]; see Corsi (2004) for further details. In general, a process obtained aggregating short memory processes can exhibit long memory features when the aggregation level is not large enough compared to the lowest frequency component of the aggregated process. Indeed, Granger (1980) shows that aggregating an infinite number of short memory processes can induce a long memory process; see also Ding and Granger (1996). Recently, LeBaron (2001) shows that the sum of only three AR(1) processes (each process operating on a different time scale) can exhibit long memory when the half life of the longest component is large enough.[5]

In our experiments, we set the parameters as follows: $\beta_0 = 0.002$, $\beta^{(d)} = 0.45$, $\beta^{(w)} = 0.30$, $\beta^{(m)} = 0.20$, and $\sigma_\omega = 0.003$. The low variance of $\omega^{(d)}$ prevents negative values of $IV^{(d)}$. As in the previous section, we simulate 10,000 sample paths of $201 \times 23{,}400$ log-returns, i.e. each sample path represents 201 days of log-returns sampled every second. The first 200 days are used to estimate the HAR-RV model (3.11) and the last day is saved for out-of-sample forecasts. The out-of-sample forecasts of IV are correctly specified using Eq. (3.11), which implies AR-type forecast equations. In-sample estimations are collected in Aït-Sahalia and Mancini (2007) and out-of-sample forecasts are reported in Table 5. These results confirm all the previous findings about the outperformance of TSRV over RV.

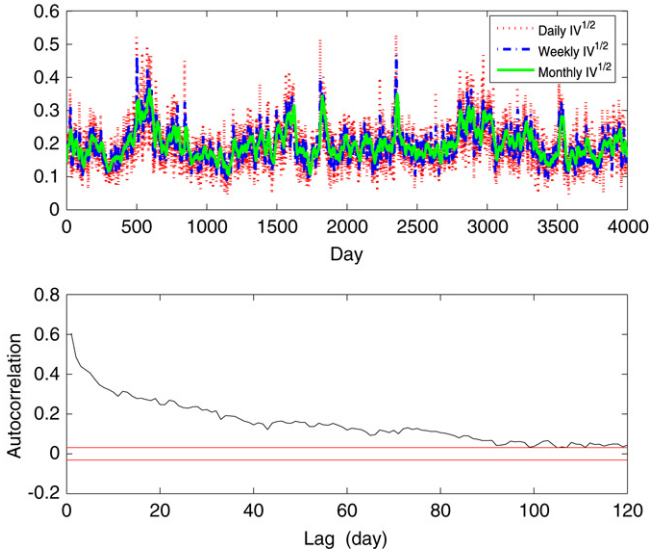

Fig. 3. First plot: simulated path of daily, weekly and monthly $IV^{1/2}$ on an annual base under the HAR-RV model, $IV^{(d)} = 0.004 + 0.35\, RV^{(d)} + 0.30\, RV^{(w)} + 0.25\, RV^{(m)} + \omega^{(d)}$, where $\omega^{(d)} \sim NID(0, 0.003^2)$. The efficient log-return $r_t = X_t - X_{t-\Delta} = \sqrt{IV_t^{(d)}} z_t$, where $z \sim NID(0, 1)$ and $\Delta = 3$ h. Second plot: estimated autocorrelation function of daily $IV^{1/2}$ with lags measured in days.

3.5. Fractional Ornstein–Uhlenbeck model

A substantial amount of empirical evidence suggests that the volatility process may exhibit long memory. Above, we have considered models which approximate this behavior. But a powerful class of models designed to capture this feature are processes driven by fractional Brownian motions. We will use a fractional Ornstein–Uhlenbeck (OU) process, as in Comte and Renault (1998), with a twist, for the purpose of representing

[4] The long sample path of 4000 days displayed in Fig. 3 is obtained by simulating $IV^{(d)}$ with time step of three hours and persistence of the process $\beta^{(d)} + \beta^{(w)} + \beta^{(m)} = 0.9$. Empirically, it has been observed that the higher the sampling frequency of returns, the more persistent is the volatility process. In our Monte Carlo experiment, we will simulate returns at frequency of one second and then we will increase the persistence of $IV^{(d)}$ to 0.95. In both cases, the unconditional integrated variance $\beta_0/(1 - \beta^{(d)} - \beta^{(w)} - \beta^{(m)}) = 0.04$.

[5] A closely related approach to modeling long-range dependence is given by the superposition of Ornstein–Uhlenbeck or CEV models; see Barndorff-Nielsen (2001) and Barndorff-Nielsen and Shephard (2001).

Table 5
Out-of-sample, one day ahead forecasts of daily IV in percentage based on 10,000 simulated paths under the HAR-RV model, $IV^{(d)} = 0.002 + 0.45\,RV^{(d)} + 0.30\,RV^{(w)} + 0.20\,RV^{(m)} + \omega^{(d)}$, where $\omega^{(d)} \sim NID(0, 0.003^2)$

	b_0	b_1	b_2	R^2
RV 5 min	−1.499 (0.017)		0.981 (0.005)	0.762
RV 5 min MA(1)	−1.162 (0.014)		0.934 (0.005)	0.804
RV 10 min	−0.727 (0.013)		0.978 (0.006)	0.757
RV 15 min	−0.466 (0.012)		0.976 (0.006)	0.734
RV 30 min	−0.111 (0.012)		0.921 (0.006)	0.670
TSRV 5 min	−0.005 (0.006)	1.017 (0.004)		0.891
TSRV 5 min + RV 5 min	−0.098 (0.017)	0.970 (0.009)	0.053 (0.009)	0.891
TSRV 5 min + RV 5 min MA(1)	−0.065 (0.016)	0.975 (0.011)	0.043 (0.011)	0.891
TSRV 5 min + RV 10 min	−0.027 (0.011)	0.996 (0.009)	0.023 (0.009)	0.891
TSRV 5 min + RV 15 min	−0.012 (0.009)	1.009 (0.008)	0.009 (0.009)	0.891
TSRV 5 min + RV 30 min	−0.017 (0.007)	0.998 (0.007)	0.023 (0.007)	0.891
TSRV 10 min	−0.015 (0.007)	1.037 (0.004)		0.860
TSRV 10 min + RV 5 min	−0.333 (0.019)	0.872 (0.010)	0.182 (0.010)	0.864
TSRV 10 min + RV 5 min MA(1)	−0.287 (0.017)	0.836 (0.013)	0.198 (0.012)	0.864
TSRV 10 min + RV 10 min	−0.118 (0.012)	0.937 (0.011)	0.109 (0.011)	0.861
TSRV 10 min + RV 15 min	−0.049 (0.010)	0.994 (0.010)	0.048 (0.011)	0.860
TSRV 10 min + RV 30 min	−0.034 (0.008)	1.003 (0.009)	0.039 (0.009)	0.860
TSRV 15 min	−0.015 (0.007)	1.051 (0.005)		0.833
TSRV 15 min + RV 5 min	−0.535 (0.019)	0.775 (0.010)	0.299 (0.010)	0.847
TSRV 15 min + RV 5 min MA(1)	−0.498 (0.017)	0.686 (0.013)	0.353 (0.012)	0.848
TSRV 15 min + RV 10 min	−0.234 (0.013)	0.831 (0.012)	0.235 (0.011)	0.840
TSRV 15 min + RV 15 min	−0.109 (0.011)	0.924 (0.012)	0.137 (0.012)	0.836
TSRV 15 min + RV 30 min	−0.045 (0.009)	0.993 (0.010)	0.064 (0.010)	0.834
TSRV 30 min	0.021 (0.009)	1.071 (0.006)		0.764
TSRV 30 min + RV 5 min MA(1)	−0.821 (0.016)	0.432 (0.012)	0.604 (0.010)	0.827
TSRV 30 min + RV 30 min	−0.072 (0.010)	0.858 (0.013)	0.220 (0.012)	0.771
TSRV minimum variance	0.009 (0.005)	0.997 (0.003)		0.918

The efficient log-return $r_t = X_t - X_{t-\Delta} = \sqrt{IV_t^{(d)}} z_t$, where $z \sim NID(0, 1)$ and $\Delta = 1$ s. The observed log-price $Y = X + \varepsilon$, where $\varepsilon \sim NID(0, 0.001^2)$. OLS standard errors in parenthesis. Eq. (3.11) is estimated using the previous 200 days of estimated daily IV, and then aggregating such estimates to obtain weekly, and monthly IV.

a process with long memory in volatility. The price dynamics continue to be given by Eq. (3.1).

Let us start with a brief description of fractional OU processes; for details, we refer the reader to Cheridito et al. (2003). A fractional Brownian motion (FBM) with Hurst parameter $H \in (0, 1]$ is an almost surely continuous, mean zero Gaussian process with covariance kernel $\mathrm{Cov}(W_{H,t}, W_{H,s}) = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |s-t|^{2H}\right)$ for all $t, s \in \mathbb{R}$. Further, $W_{H,0} = 0$ almost surely and W_H is a process with stationary increments; in general, however, the increments are not independent, except in the special case $H = 1/2$ which corresponds to a (two-sided) standard Brownian motion. When $H \neq 1/2$, W_H is not a semimartingale and for $0 < t < s$

$$\begin{aligned}
&\mathrm{Cov}(W_{H,\tau+t} - W_{H,\tau}, W_{H,\tau+s+t} - W_{H,\tau+s}) \\
&= \frac{1}{2}\left(|s-t|^{2H} + |s+t|^{2H} - 2\,|s|^{2H}\right) \\
&= \frac{s^{2H}}{2}\left(\left|1 - \frac{t}{s}\right|^{2H} + \left|1 + \frac{t}{s}\right|^{2H} - 2\right) \\
&= \sum_{n=1}^{N} \frac{t^{2n}}{(2n)!}\left(\prod_{k=0}^{2n-1}(2H-k)\right) s^{2H-2n} + O(s^{2H-2N-2}) \\
&\quad \text{as } s \to \infty
\end{aligned} \tag{3.12}$$

for every integer $N \geq 1$. Thus, for $H \in (1/2, 1]$, the increments process $(W_{H,(n+1)t} - W_{H,nt})_{n=0}^{\infty}$ exhibits long memory, namely $\sum_{n=0}^{\infty} \mathrm{Cov}(W_{H,t}, W_{H,(n+1)t} - W_{H,nt}) = \infty$. An alternative characterization of FBM is through its stochastic representation in terms of the standard BM, W:

$$\begin{aligned}
W_{H,t} = \frac{1}{\Gamma(H+1/2)}\Bigg(&\int_{-\infty}^{0}\left(|t-s|^{H-1/2} - |s|^{H-1/2}\right)\mathrm{d}W_s \\
&+ \int_0^t |t-s|^{H-1/2}\mathrm{d}W_s\Bigg),
\end{aligned} \tag{3.13}$$

where Γ is the Gamma function. The representation (3.13) is in general not unique.

A fractional OU process is the solution of

$$\mathrm{d}Z_{H,t} = \kappa(\alpha - Z_{H,t})\mathrm{d}t + \gamma\,\mathrm{d}W_{H,t}, \tag{3.14}$$

where we assume mean reversion ($\kappa > 0$) to a positive mean ($\alpha > 0$). First, there exists a stationary, almost surely continuous, solution $Z_{H,t}$ to (3.14). Second, the solution is given by

$$Z_{H,t} = \alpha + \gamma\,\mathrm{e}^{-\kappa t}\int_{-\infty}^{t}\mathrm{e}^{\kappa u}\mathrm{d}W_{H,u} \tag{3.15}$$

and is Gaussian and ergodic.[6] Third, the autocovariance function of Z_H decays at the same rate as that of W_H (see Theorem 2.3 in Cheridito et al. (2003)):

$$\begin{aligned}
\mathrm{Cov}(Z_{H,t}, Z_{H,t+s}) = \frac{\gamma^2}{2}\sum_{n=1}^{N}\kappa^{-2n}\left(\prod_{k=0}^{2n-1}(2H-k)\right)s^{2H-2n} \\
+ O(s^{2H-2N-2}) \quad \text{as } s \to \infty
\end{aligned}$$

so Z_H exhibits long memory when $H \in (1/2, 1]$.

Comte and Renault (1998) investigate the long memory properties of a volatility process driven by the fractional Ornstein–Uhlenbeck process $Z_{H,t}$, where $Z_{H,t} = \log(\sigma_t)$. But in order to be able to forecast IV exactly under the assumed model, we will modify that specification to one where $Z_{H,t} = \sigma_t$ directly. This has the drawback of allowing for negative values of σ_t, which is not pretty but otherwise creates no mathematical difficulties. As in the case of a standard OU process $Z_{1/2,t}$, a high enough long term mean α can make the probability of this occurring very small.[7]

[6] In Eq. (3.15) the integral $\int_{-\infty}^{t}\mathrm{e}^{\kappa u}\mathrm{d}W_{H,u}$ appears, where W_H is the FBM. Since the FBM is not a semimartingale, stochastic integrals with respect to FBM might not be continuous function with respect to the integrand. However, the special case of deterministic integrand turns out to be sufficient for the present purpose; see, for instance, Gripenberg and Norros (1996) for a short and elementary FBM-specific introduction to the subject.

[7] Indeed, all simulated instantaneous variances were strictly positive.

It turns out that we now have all we need to be able to compute a forecasting equation for IV. Let $I\sigma_m = \int_{T_m}^{T_{m+1}} \sigma_t \mathrm{d}t$. It is clear that $(I\sigma_m, I\sigma_{m-1})$ is jointly Gaussian. As a result,

$$E[I\sigma_m|I\sigma_{m-1}] = E[I\sigma_m] + \frac{\mathrm{Cov}(I\sigma_m, I\sigma_{m-1})}{\mathrm{Var}(I\sigma_{m-1})} \times (I\sigma_{m-1} - E[I\sigma_{m-1}]). \tag{3.16}$$

The key point is that this is an affine expression in $I\sigma_{m-1}$, hence one that is amenable to an exactly-specified forecasting regression. The coefficients of that regression expression can be computed explicitly as $E[I\sigma_m] = \alpha(T_{m+1} - T_m)$ and

$$\begin{aligned}\mathrm{Cov}(I\sigma_m, I\sigma_{m-1}) &= E\left[\left(\int_{T_m}^{T_{m+1}} Z_{H,t}\mathrm{d}t - E[I\sigma_m]\right) \times \left(\int_{T_{m-1}}^{T_m} Z_{H,r}\mathrm{d}r - E[I\sigma_{m-1}]\right)\right] \\ &= E\left[\left(\gamma \int_{T_m}^{T_{m+1}} \int_{-\infty}^{t} \mathrm{e}^{-\kappa(t-u)}\mathrm{d}W_{H,u}\mathrm{d}t\right) \times \left(\gamma \int_{T_{m-1}}^{T_m} \int_{-\infty}^{r} \mathrm{e}^{-\kappa(r-v)}\mathrm{d}W_{H,v}\mathrm{d}r\right)\right] \\ &= \gamma^2 \int_{T_m}^{T_{m+1}} \int_{T_{m-1}}^{T_m} \mathrm{e}^{-\kappa(t+r)} E\left[\left(\int_{-\infty}^{t} \mathrm{e}^{\kappa u}\mathrm{d}W_{H,u}\right) \times \left(\int_{-\infty}^{r} \mathrm{e}^{\kappa v}\mathrm{d}W_{H,v}\right)\right]\mathrm{d}r\mathrm{d}t \end{aligned} \tag{3.17}$$

and we conclude with the fact that

$$E\left[\left(\int_{-\infty}^{t} f(u)\mathrm{d}W_{H,u}\right)\left(\int_{-\infty}^{r} g(v)\mathrm{d}W_{H,v}\right)\right] = \int_{-\infty}^{t}\int_{-\infty}^{r} f(u)g(v)\,|u-v|^{2H-2}\,\mathrm{d}v\mathrm{d}u \tag{3.18}$$

(see Pipiras and Taqqu (2000)). The calculation for $\mathrm{Var}(I\sigma_{m-1})$ is identical.

In our experiments, we need to simulate relatively long sample paths of the fractional OU process. Several approaches are available to simulate fractional Gaussian processes. For instance, using a truncated version of the FBM, Comte and Renault (1998) propose a tractable representation of the fractional OU process for simulation purposes; see also Comte and Renault (1996). Unfortunately, this approach relies on a numerical evaluation of stochastic integrals and it is quite computationally demanding for simulating the long sample paths needed in the present setting.[8] For the simulation of the fractional Gaussian noise (FGN), that is the increment of FBM, we use the method proposed by Davies and Harte (1987). The aim is to compute a matrix G such that $\Sigma = GG^\top$, where $G^\top$ is the transpose of G and Σ is the covariance matrix of the FGN with autocovariance function

$$\rho(k) = \frac{\mathrm{Var}(W_{H,t+\Delta} - W_{H,t})}{2} \times \left(|k+1|^{2H} + |k-1|^{2H} - 2\,|k|^{2H}\right). \tag{3.19}$$

Suppose that a sample path of size N has to be simulated, the main idea is to embed Σ in a so-called circulant matrix C of size $2N$. The first row of such a matrix is

$$[\rho(0)\ \ \rho(1)\ \ \rho(2)\ \ \cdots\ \ \rho(N-1)\ \ 0\ \ \rho(N-1)\ \ \rho(N-2)\ \ \cdots\ \ \rho(1)]$$

and the i-th row is obtained by shifting the first row $i-1$ "places" to the right and padding the removed elements to the left. The upper left corner of size N is the covariance matrix Σ. The circulant matrix C can be decomposed as $C = Q\Lambda Q^*$, where Λ is the diagonal matrix of eigenvalues of C, Q is a unitary matrix and Q^* is the complex conjugate of $Q^\top$. Hence QQ^* is the identity matrix. A sample path of the FGN is obtained by $Q\Lambda^{1/2}Q^*\zeta$, where ζ is a vector with iid Gaussian random variables. Davies and Harte (1987) provide an algorithm in three steps to efficiently compute the last quantity; see also Craigmile (2003) and Dieker (2004) and references therein.

1. Compute the Λ eigenvalues. The eigenvalues λ_k, $k = 0, \ldots, 2N-1$, can be efficiently computed using the fast Fourier transform of the elements in the first row of C.
2. Compute $U = Q^*\zeta$. Using the covariance structure of U leads to the following simulation scheme for $U = [U_0\ \ U_1\ \ \cdots\ \ U_{N-1}\ \ U_N\ \ U_{N+1}\ \ \cdots\ \ U_{2N-1}]^\top$.
   - Generate two independent Gaussian random variables U_0 and U_N.
   - For $1 \le s \le N-1$, generate two independent Gaussian random variables $\zeta_s^{(1)}$ and $\zeta_s^{(2)}$ and
   $$U_s = \frac{1}{\sqrt{2}}(\zeta_s^{(1)} + \imath\zeta_s^{(2)}), \qquad U_{2N-s} = \frac{1}{\sqrt{2}}(\zeta_s^{(1)} - \imath\zeta_s^{(2)}), \tag{3.20}$$
   where $\imath = \sqrt{-1}$. The resulting vector U has the same distribution as $Q^*\zeta$.
3. Compute $Q\Lambda^{1/2}U$. This quantity can be efficiently computed using the inverse fast Fourier transform of the elements $\sqrt{\lambda_s}\,U_s$, $s = 0, \ldots, 2N-1$.

A sample path of the FGN is obtained by taking the first N elements of $Q\Lambda^{1/2}U$. The main advantage of the Davies–Harte method is the low computational cost which is of order $N\log(N)$ for N sample points (when $\log_2(N)$ is an integer), as opposed for instance to the Hosking method which is of order N^2. Once a sample path of the FGN is available, the fractional OU can be simulated using the Euler discretization of Eq. (3.14). We set the parameter $\alpha = 0.2, \kappa = 20, \gamma = 0.012, H = 0.7$ and we simulate 10,000 sample paths of 87 days for the fractional OU process using a time interval $\Delta = 1$ s. Hence each sample path consists of $87 \times 23{,}401$ log-prices.[9] Eq. (3.16) is estimated using the previous 86 days of IV and the last day is saved for the out-of-sample forecast of IV.

In-sample simulation results are reported in Aït-Sahalia and Mancini (2007) and confirm that TSRV largely outperforms RV in the estimation of IV. Table 6 shows the out-of-sample simulation results. Forecasts of IV given by RV are quite biased and inefficient, while TSRV forecasts are much more accurate for instance in terms of R^2. Other Monte Carlo simulations (not reported here) confirm that the levels of R^2 depend on the amount of predictability (or memory) in the data generating process. Changing model parameters will change R^2, but the outperformance of TSRV over RV and the large differences in R^2 remain.

[8] Similar computational difficulties arise for instance in the Hosking method, also known as the Durbin or Levinson method.

[9] The odd number of days allows to simulate $N = 2^{21}$ sample points using the standard "power of two" fast Fourier transform algorithm and to speed up the simulation. Since $2^{21} > 87 \times 23{,}401$, the first unnecessary sample points are discarded.

Table 6
Out-of-sample, one day ahead forecasts of daily IV in percentage based on 10,000 simulated paths under the fractional OU model, $d\sigma = 20\,(0.2 - \sigma)dt + 0.012dW_H$, where dW_H is a FBM with Hurst parameter 0.7

	b_0	b_1	b_2	R^2
RV 5 min	−1.155 (0.032)		0.873 (0.010)	0.438
RV 5 min MA(1)	−0.928 (0.026)		0.855 (0.009)	0.499
RV 10 min	−0.448 (0.024)		0.862 (0.010)	0.421
RV 15 min	−0.158 (0.022)		0.831 (0.010)	0.389
RV 30 min	0.192 (0.021)		0.760 (0.011)	0.325
TSRV 5 min	0.006 (0.009)	1.007 (0.006)		0.757
TSRV 5 min + RV 5 min	0.306 (0.024)	1.111 (0.010)	−0.147 (0.011)	0.762
TSRV 5 min + RV 5 min MA(1)	0.315 (0.021)	1.159 (0.011)	−0.185 (0.011)	0.764
TSRV 5 min + RV 10 min	0.263 (0.017)	1.150 (0.009)	−0.202 (0.011)	0.766
TSRV 5 min + RV 15 min	0.209 (0.014)	1.139 (0.009)	−0.194 (0.010)	0.766
TSRV 5 min + RV 30 min	0.159 (0.012)	1.118 (0.008)	−0.176 (0.009)	0.766
TSRV 10 min	0.035 (0.012)	1.003 (0.007)		0.645
TSRV 10 min + RV 5 min	0.017 (0.029)	0.996 (0.013)	0.009 (0.014)	0.645
TSRV 10 min + RV 5 min MA(1)	0.051 (0.026)	1.013 (0.016)	−0.011 (0.015)	0.645
TSRV 10 min + RV 10 min	0.213 (0.021)	1.129 (0.014)	−0.157 (0.015)	0.649
TSRV 10 min + RV 15 min	0.240 (0.018)	1.184 (0.014)	−0.229 (0.014)	0.654
TSRV 10 min + RV 30 min	0.207 (0.015)	1.185 (0.012)	−0.245 (0.013)	0.657
TSRV 15 min	0.080 (0.014)	0.988 (0.009)		0.567
TSRV 15 min + RV 5 min	−0.258 (0.032)	0.843 (0.015)	0.177 (0.015)	0.573
TSRV 15 min + RV 5 min MA(1)	−0.270 (0.028)	0.765 (0.018)	0.234 (0.017)	0.575
TSRV 15 min + RV 10 min	0.060 (0.023)	0.972 (0.017)	0.018 (0.017)	0.567
TSRV 15 min + RV 15 min	0.206 (0.020)	1.122 (0.017)	−0.157 (0.018)	0.570
TSRV 15 min + RV 30 min	0.244 (0.016)	1.216 (0.016)	−0.277 (0.016)	0.580
TSRV 30 min	0.210 (0.016)	0.944 (0.011)		0.432
TSRV 30 min + RV 5 min MA(1)	−0.738 (0.028)	0.302 (0.019)	0.640 (0.016)	0.511
TSRV 30 min + RV 30 min	0.248 (0.019)	1.026 (0.024)	−0.086 (0.022)	0.433
TSRV minimum variance	−0.009 (0.006)	1.010 (0.004)		0.877

The efficient log-price $dX = (0.05 - \sigma^2/2)\,dt + \sigma\,dW_1$. The observed log-price $Y = X + \varepsilon$, where $\varepsilon \sim NID(0, 0.001^2)$. Euler discretization scheme with time step $\Delta = 1$ s. OLS standard errors in parenthesis. Eq. (3.16) is estimated using the previous 86 days of estimated IV, regressing the corresponding estimates of $\int_{T_i}^{T_{i+1}} \sigma_t^2 dt$ on a constant and $\int_{T_{i-1}}^{T_i} \sigma_t^2 dt$, for $i = 1, \ldots, 85$.

3.6. Alternative correlation structures for the market microstructure noise

In the Monte Carlo experiments above, the microstructure noise term ε was assumed to be iid. In the parametric volatility context, Aït-Sahalia et al. (2005) discuss how the likelihood function is to be modified in the case of serially correlated noise and noise that is correlated with the price process. Aït-Sahalia et al. (2006) propose a version of the TSRV estimator which can deal with such serial dependence in market microstructure noise. Hansen and Lunde (2006) considered such departures from the iid noise assumption in the case of the Zhou estimator for σ time varying.

The literature reports some evidence against the iid noise assumption. For instance, Griffin and Oomen (forthcoming) provide an interesting analysis of the tick vs. transaction time sampling schemes showing that the nature of the sampling mechanism can generate distinct autocorrelogram patterns for the resulting log-returns: in transaction time the microstructure noise can be close to iid, but in tick time it is strongly autocorrelated. Hansen and Lunde (2006) argue that the microstructure noise is autocorrelated and negatively correlated with the latent return process, in particular when prices are sampled from quotations.

In this section, we remove the independent noise assumption and ε will be neither time independent nor independent of the latent price X. The underlying data generating mechanism for X is given by the Heston model (3.1) and we examine the two estimators, RV and TSRV in that context.

3.6.1. Autocorrelated microstructure noise

Following Aït-Sahalia et al. (2006), a simple model to capture the time series dependence in the microstructure noise ε is

$$\varepsilon_t = U_t + V_t, \tag{3.21}$$

where U is iid, V is an AR(1) process with first order coefficient ϱ, $|\varrho| < 1$, and $U \perp V$ are normally distributed. We set $E[U^2] = E[V^2] = 5 \times 10^{-7}$ and $\varrho = -0.2$, which imply that $(E\,\varepsilon^2)^{1/2} = 0.001$ as in the previous experiments.

In-sample estimations of IV are rather close to the ones in the iid microstructure noise case (documented in Table 1) and the corresponding results are collected in Aït-Sahalia and Mancini (2007). Hence TSRV outperforms RV in estimating the IV also under autocorrelated noise. Table 7 shows the out-of-sample forecasts of IV and largely confirms all the previous findings. For instance RV forecasts have no additional explanatory power in terms of R^2 when added to TSRV forecasts in the Mincer–Zarnowitz regressions.

3.6.2. Microstructure noise correlated with the latent price process

We simulate a microstructure noise ε not autocorrelated, but correlated with the return latent process X

$$\mathrm{Corr}(\varepsilon_t, (X_t - X_{t-\Delta})) = \vartheta \tag{3.22}$$

and we consider both positive, $\vartheta = 0.2$, and negative, $\vartheta = -0.2$, correlations. Both in- and out-of-sample results are very close to the findings in the iid noise case documented in Tables 1 and 2 and are not reported here, but collected in Aït-Sahalia and Mancini (2007).

3.6.3. Autocorrelated microstructure noise correlated with the latent price process

In this section the microstructure noise ε is autocorrelated and correlated with the latent return process, that is $\varepsilon = U + V$, where V is the same AR(1) process as in Section 3.6.1, and U is negatively correlated with the latent returns, $\mathrm{Corr}(\varepsilon_t, (X_t - X_{t-\Delta})) = -0.2$.

In- and out-of-sample simulation results are very close to the results in the autocorrelated noise case discussed in Section 3.6.1 and are reported in Aït-Sahalia and Mancini (2007). The correlation between microstructure noise and latent return process has a minor impact on the measurement and the

Table 7
Out-of-sample, one day ahead forecasts of daily IV in percentage based on 10,000 simulated paths under the Heston model, $d\sigma^2 = 5\,(0.04 - \sigma^2)dt + 0.5\,\sigma\,dW_2$

	b_0	b_1	b_2	R^2
RV 5 min	−1.476 (0.016)		0.974 (0.005)	0.813
RV 5 min MA(1)	−1.111 (0.013)		0.918 (0.004)	0.843
RV 10 min	−0.696 (0.013)		0.962 (0.005)	0.785
RV 15 min	−0.436 (0.013)		0.953 (0.005)	0.756
RV 30 min	−0.135 (0.014)		0.920 (0.006)	0.671
TSRV 5 min	0.166 (0.005)	0.993 (0.003)		0.928
TSRV 5 min + RV 5 min	0.279 (0.017)	1.047 (0.008)	−0.061 (0.009)	0.928
TSRV 5 min + RV 5 min MA(1)	0.332 (0.016)	1.101 (0.010)	−0.109 (0.010)	0.928
TSRV 5 min + RV 10 min	0.327 (0.010)	1.122 (0.008)	−0.146 (0.008)	0.930
TSRV 5 min + RV 15 min	0.273 (0.008)	1.096 (0.007)	−0.120 (0.007)	0.929
TSRV 5 min + RV 30 min	0.249 (0.007)	1.084 (0.006)	−0.114 (0.006)	0.930
TSRV 10 min	0.098 (0.007)	0.997 (0.004)		0.888
TSRV 10 min + RV 5 min	−0.148 (0.020)	0.869 (0.010)	0.139 (0.011)	0.890
TSRV 10 min + RV 5 min MA(1)	−0.078 (0.019)	0.870 (0.013)	0.124 (0.013)	0.889
TSRV 10 min + RV 10 min	0.192 (0.013)	1.084 (0.011)	−0.094 (0.011)	0.889
TSRV 10 min + RV 15 min	0.220 (0.011)	1.140 (0.010)	−0.158 (0.011)	0.891
TSRV 10 min + RV 30 min	0.206 (0.008)	1.148 (0.008)	−0.179 (0.008)	0.893
TSRV 15 min	0.076 (0.008)	1.005 (0.004)		0.853
TSRV 15 min + RV 5 min	−0.498 (0.021)	0.693 (0.011)	0.332 (0.011)	0.865
TSRV 15 min + RV 5 min MA(1)	−0.477 (0.020)	0.583 (0.014)	0.404 (0.013)	0.865
TSRV 15 min + RV 10 min	−0.027 (0.015)	0.902 (0.013)	0.109 (0.013)	0.854
TSRV 15 min + RV 15 min	0.126 (0.012)	1.072 (0.013)	−0.071 (0.013)	0.853
TSRV 15 min + RV 30 min	0.197 (0.010)	1.209 (0.010)	−0.228 (0.011)	0.859
TSRV 30 min	0.096 (0.010)	1.014 (0.006)		0.759
TSRV 30 min + RV 5 min MA(1)	−0.982 (0.017)	0.147 (0.012)	0.801 (0.011)	0.845
TSRV 30 min + RV 30 min	0.109 (0.013)	1.045 (0.017)	−0.032 (0.017)	0.759
TSRV minimum variance	0.331 (0.004)	1.390 (0.003)		0.941

The efficient log-price $dX = (0.05 - \sigma^2/2)dt + \sigma\,dW_1$, and correlation $\rho = -0.5$ between Brownian motions. The observed log-price $Y = X + \varepsilon$, where $\varepsilon = U + V$, U is iid, V is AR(1) with first order coefficient $\varrho = -0.2$, $U \perp V$, and $E\,\varepsilon^2 = 0.001^2$. Euler discretization scheme with time step $\Delta = 1$ s. OLS standard errors in parenthesis. Eq. (3.6) is estimated using the previous 100 simulated days and regressing the corresponding estimates of $\int_{T_i}^{T_{i+1}} \sigma_t^2 dt$ on a constant and $\int_{T_{i-1}}^{T_i} \sigma_t^2 dt$, for $i = 1, \ldots, 99$.

forecast of the IV, while the autocorrelation in the microstructure noise plays a major role. Overall, TSRV outperforms RV in estimating and forecasting IV also when the microstructure noise is autocorrelated and correlated with the latent price process.

3.7. Comparison between TSRV and MSRV

We have seen that TSRV provides the first consistent and asymptotic (mixed) normal estimator of IV and has the rate of convergence $n^{-1/6}$. At the cost of higher complexity TSRV can be generalized to multiple time scales, by averaging not on two time scales but on multiple time scales. The resulting estimator, multiple scale realized volatility (MSRV), is

$$\widehat{\langle X, X\rangle}_T^{(\mathrm{msrv})} = \underbrace{\sum_{i=1}^{\mathcal{M}} a_i\,[Y, Y]_T^{(K_i)}}_{\text{weighted sum of } \mathcal{M} \text{ slow time scales}} + 2\,\underbrace{\widehat{E[\varepsilon^2]}}_{\text{fast time scale}}, \quad (3.23)$$

where $\widehat{E[\varepsilon^2]}$ is given in (2.2). TSRV corresponds to the special case $\mathcal{M} = 1$ and uses only one slow time scale (and the fast time scale to bias correct it). For suitably selected weights a_i's, $\widehat{\langle X, X\rangle}_T^{(\mathrm{msrv})}$ converges to the true $\langle X, X\rangle_T$ at the rate $n^{-1/4}$. This generalization is due to Zhang (2006).

We repeat all the previous Monte Carlo experiments to compare in-sample estimations and out-of-sample forecasts of TSRV and MSRV. To save space we report all the details and simulation results in Aït-Sahalia and Mancini (2007). Overall, under all scenarios TSRV and MSRV provide rather close (and accurate) estimations and forecasts of IV. The efficiency loss in using TSRV instead of MSRV is essentially negligible. These results are in agreement with the empirical findings of Aït-Sahalia et al. (2006) as they show that TSRV and MSRV give rather close estimates of IV for the Intel and Microsoft stocks.

4. Empirical analysis

4.1. The data

Our empirical analysis is based on the transaction prices from the NYSE's Trade and Quote (TAQ) database for the thirty DJIA stocks. The sample period extends from January 3, 2000 until December 31, 2004, for a total of 1256 trading days and the daily transaction records extend from 9:30 until 16:00. The DJIA stocks are among the most actively traded US equities and the sample period covers high and low volatility periods. Hence the comparisons between TSRV and RV and in particular the out-of-sample forecasts of the integrated volatilities will be interesting exercises.

The original data are pre-processed eliminating obvious data errors (such as transaction prices reported at zero, transaction times that are out of order, etc.) and "bounce back" outliers larger than the cutoff 0.03. Admittedly, other cutoff thresholds could be conceived, but they would be all equally arbitrary. Aït-Sahalia et al. (2006) investigate the dependence of TSRV and RV on cutoff levels, and they show empirically that TSRV is much less affected than RV by the different level choices. Such outliers can be viewed as a form of microstructure noise and the robustness of TSRV to different ways of pre-processing the data is a desirable property. The pre-processed data will be used to estimate the daily integrated variances applying the TSRV and RV estimators.

4.2. Unconditional return and integrated volatility distributions

In this section we present the statistical properties of the unconditional return and the integrated variance distributions. For each stock, daily integrated variances are estimated using the TSRV estimator (2.5) with a slow time scale of five minutes and two RV estimators based on de-meaned MA(1)-filtered and unfiltered 5 min log-returns. In the construction of the RV estimators we

Table 8
Summary statistics for the daily log-return distributions in percentage of the DJIA stocks, $r_{i,t}$

Stock	$r_{i,t}$			
	Mean	Std.	Skew.	Kurt.
Min.	−0.132	1.363	−0.612	4.105
0.10	−0.101	1.473	−0.455	4.405
0.25	−0.049	1.757	0.049	4.862
0.50	−0.011	1.883	0.205	5.477
0.75	0.056	2.064	0.322	8.386
0.90	0.071	2.294	0.411	9.418
Max.	0.118	2.787	0.759	12.277
Mean	−0.007	1.946	0.097	6.990
Std.	0.086	0.454	0.449	2.869

The sample period extends from January 3, 2000 to December 31, 2004, for a total of 1256 observations.

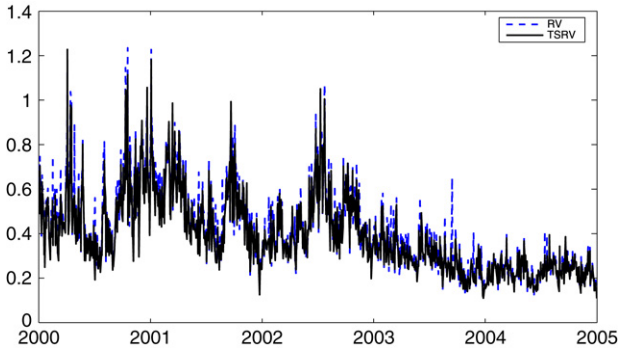

Fig. 4. Integrated volatility, $IV^{1/2}$, for Intel stock (ticker INTC) on an annual base (252 days) given by TSRV and RV from January 3, 2000 to December 31, 2004.

follow Andersen et al. (2001). The five minutes log-return series are obtained by taking the difference between log-prices recorded at or immediately before the corresponding 5 min marks. Then to compute the first RV estimator, the raw log-returns are de-meaned and for each stock an MA(1) model is estimated using the 5 min log-return series of the full five years sample. The second RV estimator is computed using the unfiltered 5 min log-returns. In our sample, the IV estimates given by the two RV estimators are very similar, confirming the findings in Andersen et al. (2001). For all stocks and all trading days the differences between the corresponding annualized integrated volatilities are always below 1%. Therefore, to save space the subsequent analysis will only involve the RV estimates based on 5 min de-meaned MA(1) filtered returns. This choice is consistent with the previous Monte Carlo analysis, where the RV forecasts based on MA(1) filtered returns tends to outperform the corresponding forecasts based on unfiltered returns. For completeness, we report all the summary statistics for both RV estimators for each DJIA stock in Aït-Sahalia and Mancini (2007).

Table 8 presents summary statistics for the daily log-returns of the DJIA stocks, $r_{i,t}$, $i = 1, \ldots, 30$, over the sample period, $t = 1, \ldots, 1256$, i.e. January 3, 2000 to December 31, 2004. The average returns across all stocks are very close to zero, the majority of the stock returns have positive skewness, and all stock returns display excess kurtosis. These findings are broadly consistent with the large empirical literature on speculative asset returns, dating back at least to Mandelbrot (1963) and Fama (1965). As observed for instance in Andersen et al. (2001), when the log-returns are standardized by the corresponding integrated volatility, $r_{i,t}/IV_{i,t}^{1/2}$, the unconditional distributions are very close to the Gaussian distribution. Table 9 shows the previous summary statistics for the log-returns standardized using TSRV and RV and confirms this finding. For instance the average kurtosis is reduced from 6.990 for the raw log-returns $r_{i,t}$ to 2.886 and 2.796 for the log-returns standardized using TSRV and RV, respectively. We remark that the integrated volatilities for standardizing the returns are only observable ex-post, and not ex-ante as for instance in GARCH-type models. Nonetheless, the approximate Gaussian distribution of standardized returns, $r_{i,t}/IV_{i,t}^{1/2}$, is in contrast to the typical finding that, when standardizing daily returns by the one day ahead GARCH-type volatility forecasts, the resulting distributions are invariably leptokurtic, albeit less so than the raw returns, $r_{i,t}$.

Table 10 summarizes the unconditional distribution of the integrated volatilities, $IV^{1/2}$, on an annual base (252 days), given by TSRV and RV for the DJIA stocks. On average, the integrated volatilities estimated using TSRV are lower, and less volatile, positively skewed and leptokurtic than the integrated volatilities given by RV. These findings are consistent with the fact that the RV estimates can be adversely affected by microstructure noise as demonstrated in the Monte Carlo analysis. In particular, the $IV^{1/2}$ estimated using RV are 2% higher with a kurtosis 23% larger than the $IV^{1/2}$ estimated using TSRV. These differences are economically significant, resulting in different asset prices and optimal asset allocations. For example, with a typical three-month option with a vega sensitivity to volatility of 10, such a 2% difference in volatility results in an option price that is 20% higher. To illustrate typical differences recorded on a single stock, Fig. 4 shows the integrated volatilities of the Intel stock (ticker INTC) estimated using TSRV and RV: as can be seen, in several occasions the two estimates are quite different.

As discussed earlier, given the high skewness and kurtosis of the integrated volatilities, a logarithmic transformation of $IV^{1/2}$ is sometimes adopted to induce an approximate Gaussian distribution. Table 11 shows the summary statistics for the $\log(IV_{i,t}^{1/2})$ distributions estimated using TSRV and RV. Both distributions are closer to the Gaussian one when compared to the unconditional distributions of $IV^{1/2}$. For instance, the average kurtosis across all the thirty stocks are now reduced to 2.987 and 2.900 for TSRV and RV estimates, compared to 9.856 and 12.074 for the $IV^{1/2}$ in Table 10. As observed in Andersen et al. (2003), the approximate normality of the log-integrated volatility, $\log(IV^{1/2})$, suggests the use of standard linear Gaussian approaches for modeling and forecasting logarithmic volatility. In our in- and out-of-sample forecasts of the DJIA integrated volatilities, we will use AR models for $\log(IV^{1/2})$ to obtain $IV^{1/2}$ forecasts. We will also experiment with AR models for the integrated variance processes, IV, to obtain integrated volatility forecasts.

4.3. Temporal dependence and long memory of integrated volatility

Temporal dependence and long memory are well documented empirical features of financial asset returns. Table 12 reports summary statistics to investigate the temporal dependence of the integrated volatility, $IV_{i,t}^{1/2}$, in Panel A, and the logarithmic integrated volatility, $\log(IV_{i,t}^{1/2})$, in Panel B, estimated using TSRV and RV. Consistently with the empirical literature, all the Ljung–Box portmanteau test strongly rejects the joint null hypothesis of zero autocorrelations up to lag 22, that is about one month of trading days. All the test statistics are well above the 40.289 critical value at 1% confidence level, documenting the well-known volatility clustering phenomenon. Similar slow decay rates in the autocorrelation functions have been documented in the empirical literature with daily time series of absolute and squared returns spanning several decades (see, for example, Ding et al. (1993)), but the results in Table 12 are noteworthy in that the sample only spans five years. In spite of the slow decay in the autocorrelations, about 90% of the augmented Dickey–Fuller tests

Table 9
Summary statistics for the standardized log-return distributions of the DJIA stocks, $r_{i,t}/IV_{i,t}^{1/2}$

Stock	TSRV $r_{i,t}/IV_{i,t}^{1/2}$				RV $r_{i,t}/IV_{i,t}^{1/2}$			
	Mean	Std.	Skew.	Kurt.	Mean	Std.	Skew.	Kurt.
Min.	−0.098	0.927	−0.080	2.564	−0.091	0.864	−0.002	2.582
0.10	−0.041	0.936	−0.011	2.644	−0.043	0.878	0.018	2.614
0.25	−0.031	0.956	0.018	2.737	−0.027	0.895	0.043	2.667
0.50	0.005	0.993	0.106	2.791	0.006	0.937	0.101	2.774
0.75	0.032	1.026	0.154	2.931	0.032	0.962	0.140	2.842
0.90	0.067	1.053	0.197	3.062	0.061	0.986	0.162	2.989
Max.	0.103	1.092	0.266	3.475	0.105	1.019	0.287	3.107
Mean	0.006	0.998	0.093	2.886	0.006	0.934	0.107	2.796
Std.	0.064	0.058	0.114	0.286	0.061	0.054	0.093	0.182

The sample extends from January 3, 2000 to December 31, 2004, for a total of 1256 observations. TSRV is computed using a slow time scale of 5 min. RV is computed using MA(1) filtered 5 min log-returns, as detailed in the text.

Table 10
Summary statistics for the volatility distributions of the DJIA stocks on annual base (252 days), $IV_{i,t}^{1/2}$

Stock	TSRV $IV_{i,t}^{1/2}$				RV $IV_{i,t}^{1/2}$			
	Mean	Std.	Skew.	Kurt.	Mean	Std.	Skew.	Kurt.
Min.	0.210	0.092	1.069	4.376	0.223	0.100	0.980	3.928
0.10	0.218	0.098	1.241	5.217	0.231	0.107	1.258	5.047
0.25	0.248	0.108	1.394	6.040	0.269	0.121	1.373	6.002
0.50	0.275	0.124	1.506	7.106	0.297	0.133	1.660	8.379
0.75	0.297	0.136	1.903	10.534	0.319	0.154	2.110	13.762
0.90	0.319	0.157	2.155	15.004	0.347	0.173	2.530	18.949
Max.	0.394	0.180	2.981	20.713	0.415	0.195	3.072	28.451
Mean	0.280	0.128	1.750	9.856	0.300	0.141	1.855	12.074
Std.	0.059	0.030	0.611	5.587	0.063	0.032	0.696	8.320

The sample extends from January 3, 2000 to December 31, 2004, for a total of 1256 observations. TSRV is computed using a slow time scale of 5 min. RV is compute using MA(1) filtered 5 min log-returns, as detailed in the text.

Table 11
Summary statistics for the log-volatility distributions of the DJIA stocks in percentage on daily base, $\log(IV_{i,t}^{1/2})$

Stock	TSRV $\log(IV_{i,t}^{1/2})$				RV $\log(IV_{i,t}^{1/2})$			
	Mean	Std.	Skew.	Kurt.	Mean	Std.	Skew.	Kurt.
Min.	0.176	0.342	−0.300	2.490	0.229	0.358	−0.265	2.424
0.10	0.234	0.369	−0.083	2.578	0.288	0.392	−0.032	2.513
0.25	0.335	0.389	0.062	2.761	0.405	0.406	0.006	2.717
0.50	0.456	0.418	0.154	2.913	0.514	0.433	0.142	2.872
0.75	0.535	0.447	0.255	3.117	0.600	0.460	0.266	3.073
0.90	0.625	0.475	0.332	3.323	0.689	0.489	0.328	3.253
Max.	0.819	0.551	0.460	3.725	0.865	0.552	0.431	3.447
Mean	0.454	0.427	0.126	2.987	0.513	0.442	0.125	2.900
Std.	0.210	0.066	0.239	0.405	0.209	0.060	0.222	0.351

The sample extends from January 3, 2000 to December 31, 2004, for a total of 1256 observations. TSRV is computed using a slow time scale of 5 min. RV is compute using MA(1) filtered 5 min log-returns, as detailed in the text.

for $IV_{i,t}^{1/2}$ and $\log(IV_{i,t}^{1/2})$ estimated using TSRV and RV reject the null hypothesis of unit root, (H_0 : ADF = 0, in Table 12), when judged by the conventional −3.432 critical value at 5% confidence level. It is known, however, that the outcome of standard unit root tests should be carefully interpreted with slowly decaying memory processes; see, for instance, Schwert (1987).

Following the theoretical contribution of Robinson (1991), many subsequent studies suggested the empirical relevance of long memory in asset return volatility, including for instance, Ding et al. (1993), Baillie et al. (1996) and Breidt et al. (1998).[10] Table 12 shows the degrees of fractional integration, d_{GPH}, obtained using the Geweke and Porter-Hudak (1983) log-periodogram regression estimator as formally developed by Robinson (1995), for the integrated volatility and the logarithmic integrated volatility estimated using TSRV and RV. For $IV_{i,t}^{1/2}$ and $\log(IV_{i,t}^{1/2})$, respectively in five and six cases, TSRV gives point estimates of d_{GPH} above 0.5, although all these estimates are not statistically different from 0.465 at 1% confidence level. For all stocks and all integrated volatilities and logarithmic integrated volatilities the null hypothesis of no long memory (H_0 : $d_{\mathrm{GPH}} = 0$) is strongly rejected. The estimates of the fractional parameter d_{GPH} have to be carefully interpreted because of the measurement error in $IV_{i,t}^{1/2}$.

Table 12 shows that the Ljung–Box statistics and the degree of fractional integrations are almost always larger for TSRV than RV. These findings are consistent with the fact that TSRV provides more precise and less noisy estimates of the integrated volatility process. In contrast, RV can be largely affected by the microstructure noise and this noise masks the strong persistence in the integrated volatility process.

[10] Long memory, or fractionally integrated volatility processes also help to explain empirical features of options data, such as the persistent volatility smiles even for long maturity options; see, for example, Comte and Renault (1998) and Bollerslev and Mikkelsen (1996).

Table 12
Summary statistics of the time series dependence for the DJIA stocks integrated volatility, $IV_{i,t}^{1/2}$, in Panel A, and log-integrated volatility, $\log(IV_{i,t}^{1/2})$, in Panel B

Stock	TSRV $IV_{i,t}^{1/2}$			RV $IV_{i,t}^{1/2}$		
	Q_{22}	ADF	d_{GPH}	Q_{22}	ADF	d_{GPH}
Panel A						
Min.	2,998.953	−6.191	0.269	2,881.977	−6.132	0.290
0.10	4,753.831	−5.211	0.426	4,632.983	−5.281	0.351
0.25	7,213.061	−4.755	0.433	6,310.282	−4.745	0.381
0.50	8,704.990	−4.511	0.460	7,205.355	−4.350	0.406
0.75	10,684.396	−4.145	0.487	9,824.078	−4.161	0.435
0.90	11,803.375	−3.848	0.502	11,258.890	−4.000	0.471
Max.	12,567.612	−2.662	0.516	11,709.776	−2.686	0.489
Mean	8,389.460	−4.475	0.442	7,689.049	−4.479	0.403
Std.	3,341.030	1.025	0.077	3,117.939	1.001	0.064
Panel B						
Min.	4,132.995	−5.492	0.357	4,038.436	−5.683	0.355
0.10	7,152.367	−4.658	0.406	7,372.738	−4.760	0.368
0.25	8,683.572	−4.352	0.425	8,416.448	−4.443	0.388
0.50	11,562.323	−4.051	0.440	10,676.330	−4.055	0.413
0.75	13,624.768	−3.741	0.475	13,157.981	−3.770	0.451
0.90	14,658.166	−3.429	0.519	13,582.058	−3.488	0.461
Max.	16,293.398	−2.341	0.528	15,705.337	−2.453	0.479
Mean	10,872.513	−4.009	0.450	10,421.333	−4.093	0.416
Std.	4,067.642	0.920	0.057	3,770.563	0.946	0.045

The sample extends from January 3, 2000 to December 31, 2004, for a total of 1256 observations. TSRV is computed using a slow time scale of 5 min. RV is compute using MA(1) filtered 5 min log-returns, as detailed in the text. The table reports the Ljung–Box portmanteau test for up to 22nd order autocorrelation, Q_{22}, the 1% critical value is 40.289. The augmented Dickey–Fuller test for unit root involving 22 augmentation lags, a constant term and time trend, ADF, the 5% critical value is −3.432, and at 1% critical value is −3.998. The Geweke and Porter-Hudak estimate for the degree of fractional integration, d_{GPH}, the estimates are based on the first $m = 1256^{9/10} = 615$ Fourier frequencies, the asymptotic standard error for all of the d_{GPH} estimates is $\pi(24m)^{-1/2} = 0.026$.

4.4. Forecasts of the integrated volatility

In this section we investigate the in- and out-of-sample forecasts of the integrated volatility given by TSRV and RV. We use two approaches to forecast $IV^{1/2}$. In the first approach we use an AR(1) model with constant for IV and in the second approach an AR(1) model with constant for the logarithmic integrated volatility, $\log(IV^{1/2})$. For each stock the AR models are estimated on a moving window of 100 trading days. As in the Monte Carlo analysis, the in-sample forecasts are obtained for the 100th day and the out-of-sample forecasts for the 101st day. Hence in- and out-of-sample forecast results can be consistently compared. All the AR models are separately estimated for both time series of integrated volatilities given by TSRV and RV. Of course, we do not expect that a simple AR(1) model perfectly captures the dynamic of the integrated volatility. This model is mainly chosen to make the empirical analysis comparable with the Monte Carlo analysis. Nevertheless, the moving window estimation of AR models will deliver relatively high R^2 for the Mincer–Zarnowitz forecast evaluation regressions, that are for instance higher than the ones reported in Andersen et al. (2003), meaning that overall the volatility forecasts are reasonably good.

Moreover, as we already mentioned, the log-normality of the integrated volatility documented in Table 11 suggests the use of linear models to describe the integrated volatility dynamics. To evaluate the volatility forecasts, some benchmark integrated volatility has to be assumed, as in empirical applications the true integrated volatility is unknown. The standard practice is to use the best available estimate as the true integrated volatility.[11] In our setting, theoretical considerations and the previous Monte Carlo analysis suggest to use TSRV as a proxy for the true integrated variance.

Tables 13 and 14 show, respectively, in-sample and out-of-sample Mincer–Zarnowitz forecast evaluation regression (3.7) for the integrated volatility, $IV^{1/2}$, using AR models for $\log(IV^{1/2})$. The corresponding forecast results based on AR models for IV are quite similar and are omitted here, but collected in Aït-Sahalia and Mancini (2007). The Newey and West (1987) robust standard errors are reported in parenthesis.[12] In terms of R^2, TSRV forecasts are more accurate than RV forecasts. The regression coefficients b_1 associated to the TSRV forecasts are always larger than the coefficients b_2 associated to the RV forecasts, except in one occasion for the Altria Group Inc. stock (ticker MO); see Aït-Sahalia and Mancini (2007) for details. Moreover, forecasts based on the RV estimator have no additional explanatory power when compared to TSRV forecasts.

When compared to the in-sample forecasts, in terms of R^2 the deterioration of the out-of-sample forecasts is quite modest; see also Aït-Sahalia and Mancini (2007) for in- and out-of-sample forecasts on a stock by stock base. The moving window estimation of AR models, the strong persistence of the integrated volatility process, and the one day ahead horizon for the out-of-sample forecasts explain this result. Certainly integrated volatilities can be forecasted for longer time horizons, such as a week, and in those cases the deterioration in the forecasting performances would be more pronounced. However, the ranking of the out-of-sample forecast performances will be hardly reverted. Our goal is to compare the TSRV and RV forecasts, and even considering longer time horizons it is quite reasonable to believe that TSRV will continue to outperform RV.

Finally, the overall conclusion of the empirical analysis is that both in- and out-of-sample forecasts on real data largely confirm the previous simulation results.

5. Conclusions

We compared in-sample estimations and out-of-sample forecasts of quadratic variation given by the two scales realized

[11] For instance, Andersen et al. (2003) and Corsi (2004) use the RV estimates as true integrated variances.

[12] In the estimation of the robust covariance matrix we set the number of lags equal to 20, as most autocorrelations of the residuals in the Mincer–Zarnowitz regressions are not significantly different from zero beyond that lag.

Table 13
In-sample one day ahead forecasts of $IV^{1/2}$ in percentage using an AR(1) model with constant for $\log(IV^{1/2})$ for the DJIA stocks

	Stock	b_0	b_1	b_2	R^2
RV 5 min MA(1)	Min.	−0.098 (0.050)		0.759 (0.031)	0.353
	0.10	−0.092 (0.064)		0.861 (0.044)	0.418
	0.25	−0.041 (0.067)		0.925 (0.049)	0.501
	0.50	−0.004 (0.080)		0.960 (0.055)	0.578
	0.75	0.037 (0.095)		0.980 (0.064)	0.640
	0.90	0.122 (0.105)		1.006 (0.070)	0.661
	Max.	0.255 (0.113)		1.022 (0.082)	0.709
	Mean	0.025 (0.082)		0.930 (0.056)	0.551
	Std.	0.118 (0.022)		0.086 (0.016)	0.122
TSRV 5 min	Min.	−0.099 (0.039)	1.000 (0.029)		0.380
	0.10	−0.086 (0.051)	1.008 (0.037)		0.451
	0.25	−0.064 (0.057)	1.019 (0.043)		0.542
	0.50	−0.048 (0.066)	1.035 (0.052)		0.595
	0.75	−0.018 (0.085)	1.054 (0.057)		0.668
	0.90	−0.004 (0.092)	1.060 (0.062)		0.686
	Max.	0.030 (0.097)	1.065 (0.076)		0.723
	Mean	−0.041 (0.070)	1.034 (0.051)		0.578
	Std.	0.043 (0.021)	0.024 (0.015)		0.118
Both	Min.	−0.120 (0.041)	0.523 (0.094)	−0.333 (0.086)	0.380
	0.10	−0.083 (0.049)	0.806 (0.131)	−0.267 (0.119)	0.452
	0.25	−0.066 (0.057)	0.964 (0.167)	−0.130 (0.143)	0.542
	0.50	−0.038 (0.065)	1.120 (0.182)	−0.077 (0.170)	0.596
	0.75	−0.011 (0.083)	1.161 (0.214)	0.052 (0.183)	0.669
	0.90	0.001 (0.090)	1.299 (0.230)	0.223 (0.211)	0.686
	Max.	0.046 (0.095)	1.390 (0.337)	0.512 (0.302)	0.725
	Mean	−0.039 (0.068)	1.037 (0.193)	−0.003 (0.173)	0.579
	Std.	0.052 (0.020)	0.277 (0.073)	0.272 (0.065)	0.118

AR(1) models are estimated on a rolling window of 100 days, from January 3, 2000 to December 31, 2004. Robust standard errors in parenthesis.

Table 14
Out-of-sample one day ahead forecasts of $IV^{1/2}$ using an AR(1) model with constant for $\log(IV^{1/2})$ for the DJIA stocks

	Stock	b_0	b_1	b_2	R^2
RV 5 min MA(1)	Min.	−0.068 (0.053)		0.735 (0.032)	0.332
	0.10	−0.050 (0.062)		0.829 (0.042)	0.381
	0.25	−0.007 (0.069)		0.907 (0.048)	0.477
	0.50	0.033 (0.082)		0.936 (0.056)	0.545
	0.75	0.083 (0.101)		0.961 (0.066)	0.617
	0.90	0.173 (0.110)		0.980 (0.075)	0.641
	Max.	0.293 (0.135)		1.000 (0.089)	0.696
	Mean	0.065 (0.087)		0.907 (0.058)	0.527
	Std.	0.120 (0.027)		0.087 (0.018)	0.126
TSRV 5 min	Min.	−0.067 (0.039)	0.959 (0.030)		0.353
	0.10	−0.043 (0.051)	0.984 (0.036)		0.421
	0.25	−0.026 (0.057)	1.005 (0.045)		0.513
	0.50	−0.013 (0.066)	1.016 (0.053)		0.576
	0.75	0.008 (0.087)	1.029 (0.058)		0.642
	0.90	0.036 (0.096)	1.036 (0.063)		0.670
	Max.	0.086 (0.103)	1.048 (0.080)		0.709
	Mean	−0.003 (0.071)	1.011 (0.052)		0.555
	Std.	0.048 (0.022)	0.029 (0.016)		0.123
Both	Min.	−0.071 (0.041)	0.409 (0.069)	−0.339 (0.046)	0.355
	0.10	−0.052 (0.050)	0.753 (0.128)	−0.243 (0.114)	0.421
	0.25	−0.027 (0.058)	0.945 (0.162)	−0.106 (0.135)	0.515
	0.50	−0.009 (0.065)	1.092 (0.190)	−0.069 (0.169)	0.576
	0.75	0.012 (0.084)	1.131 (0.208)	0.057 (0.187)	0.642
	0.90	0.042 (0.093)	1.258 (0.225)	0.237 (0.210)	0.671
	Max.	0.109 (0.102)	1.375 (0.343)	0.593 (0.311)	0.711
	Mean	0.000 (0.070)	0.995 (0.189)	0.019 (0.167)	0.556
	Std.	0.057 (0.021)	0.304 (0.079)	0.293 (0.077)	0.123

AR(1) models are estimated on a rolling window of 100 days, from January 3, 2000 to December 31, 2004. Robust standard errors in parenthesis.

volatility estimator and the realized volatility estimator, under five different dynamics for the volatility process. Overall, the different volatility models account for well-known empirical stylized facts of volatility processes, such as volatility clustering and long memory features. In all Monte Carlo experiments the TSRV estimator largely outperforms the RV estimator. An empirical application to the DJIA stocks confirms the simulation results.

References

Aït-Sahalia, Y., Mancini, L., 2007. Out of sample forecasts of quadratic variation: Appendix. Tech. Rep. Princeton University. http://www.princeton.edu/~yacine/research.htm.

Aït-Sahalia, Y., Mykland, P.A., Zhang, L., 2005. How often to sample a continuous-time process in the presence of market microstructure noise. Review of Financial Studies 18, 351–416.

Aït-Sahalia, Y., Mykland, P.A., Zhang, L., 2006. Ultra high frequency volatility estimation with dependent microstructure noise. Tech. Rep. Princeton University.

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., Diebold, F.X., Ebens, H., 2001. The distribution of realized stock return volatility. Journal of Financial Economics 61, 43–76.

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

Andersen, T.G., Bollerslev, T., Meddahi, N., 2004. Analytic evaluation of volatility forecasts. International Economic Review 45, 1079–1110.

Andersen, T.G., Bollerslev, T., Meddahi, N., 2005. Correcting the errors: Volatility forecast evaluation using high frequency data and realized volatilities. Econometrica 73, 279–296.

Andersen, T.G., Bollerslev, T., Meddahi, N., 2006. Market microstructure noise and realized volatility forecasting. Tech. Rep. University of Montreal.

Baillie, R.T., Bollerslev, T., Mikkelsen, H.O., 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 74, 3–30.

Bandi, F.M., Russell, J.R., 2006. Separating microstructure noise from volatility. Journal of Financial Economics 79, 655–692.

Barndorff-Nielsen, O.E., 2001. Superposition of Ornstein-Uhlenbeck type processes. Theory of Probability and its Applications 45, 175–194.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2006. Regular and modified kernel-based estimators of integrated variance: The case with independent noise. Tech. Rep. Department of Mathematical Sciences, University of Aarhus.

Barndorff-Nielsen, O.E., Shephard, N., 2001. Non-gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics (with discussion). Journal of the Royal Statistical Society, B 63, 167–241.

Barndorff-Nielsen, O.E., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society, B 64, 253–280.

Bollerslev, T., Mikkelsen, H.O., 1996. Modeling and pricing long memory in stock market volatility. Journal of Econometrics 73, 151–184.

Bollerslev, T., Zhou, H., 2002. Estimating stochastic volatility diffusions using conditional moments of integrated volatility. Journal of Econometrics 109, 33–65.

Breidt, F.J., Crato, N., de Lima, P., 1998. The detection and estimation of long memory in stochastic volatility. Journal of Econometrics 83, 325–348.

Cheridito, P., Kawaguchi, H., Maejima, M., 2003. Fractional Ornstein-Uhlenbeck processes. Electronic Journal of Probability 8, 1–14.

Chong, Y.Y., Hendry, D.F., 1986. Econometric evaluation of linear macroeconomic models. Review of Economic Studies 53, 671–690.

Comte, F., Renault, E., 1996. Long memory continuous time models. Journal of Econometrics 73, 101–149.

Comte, F., Renault, E., 1998. Long memory in continuous-time stochastic volatility models. Mathematical Finance 8, 291–323.

Corradi, V., Distaso, W., Swanson, N.R., 2005. Predictive inference for integrated volatility. Tech. Rep. Rutgers University.

Corradi, V., Distaso, W., Swanson, N.R., 2008. Predictive density estimators for daily volatility based on the use of realized measures. Journal of Econometrics (forthcoming).

Corsi, F., 2004. A simple long memory model of realized volatility. Tech. Rep. University of Lugano.

Cox, J.C., Ingersoll, J.E., Ross, S.A., 1985. A theory of the term structure of interest rates. Econometrica 53, 385–408.

Craigmile, P.F., 2003. Simulating a class of stationary Gaussian processes using the Davies–Harte algorithm, with application to long memory processes. Journal of Time Series Analysis 24, 505–511.

Dacorogna, M., Müller, U.A., Olsen, R.B., Pictet, O.V., 1998. Modelling short-term volatility with GARCH and HARCH models. In: Dunis, C.L., Zhou, B. (Eds.), Nonlinear modelling of high frequency financial time series. John Wiley & Sons.

Davies, R.B., Harte, D.S., 1987. Tests for the Hurst effect. Biometrika 74, 95–102.

Dieker, T., 2004. Simulation of fractional Brownian motion. Tech. Rep. University of Twente.

Ding, Z., Granger, C., 1996. Modeling volatility persistence of speculative returns: A new approach. Journal of Econometrics 73, 185–215.

Ding, Z., Granger, C.W., Engle, R.F., 1993. A long memory property of stock market returns and a new model markets. Journal of Empirical Finance 1, 83–106.

Fama, E.F., 1965. The behavior of stock market prices. Journal of Business 38, 34–105.

French, K.R., Schwert, G.W., Stambaugh, R.F., 1987. Expected stock returns and volatility. Journal of Financial Economics 19, 3–29.

Garcia, R., Meddahi, N., 2006. Comments on "Realized variance and market microstructure noise". Journal of Business and Economic Statistics 24, 184–192.

Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. Journal of Time Series Analysis 4, 221–238.

Ghysels, E., Sinko, A., 2006a. Comments on "Realized variance and market microstructure noise". Journal of Business and Economic Statistics 24, 192–194.

Ghysels, E., Sinko, A., 2006b. Volatility forecasting and microstructure noise. Tech. Rep. University of North Carolina.

Ghysels, E., Sinko, A., Valkanov, R., 2008. MIDAS regressions: Further results and new directions. Econometric Reviews (forthcoming).

Goncalves, S., Meddahi, N., 2005a. Bootstrapping realized volatility. Tech. Rep. Université de Montréal.

Goncalves, S., Meddahi, N., 2005b. Box–Cox transforms for realized volatility. Tech. Rep. Université de Montréal.

Granger, C., 1980. Long memory relationships and the aggregation of dynamic models. Journal of Econometrics 14, 227–238.

Griffin, J.E., Oomen, R.C., 2008. Sampling returns for realized variance calculations: Tick time or transaction time? Econometric Reviews (forthcoming).

Gripenberg, G., Norros, I., 1996. On the prediction of fractional Brownian motion. Journal of Applied Probability 33, 400–410.

Hansen, P.R., Lunde, A., 2006. Realized variance and market microstructure noise. Journal of Business and Economic Statistics 24, 127–161.

Heston, S., 1993. A closed-form solution for options with stochastic volatility with applications to bonds and currency options. Review of Financial Studies 6, 327–343.

Jacod, J., 1994. Limit of random measures associated with the increments of a Brownian semimartingale. Tech. Rep. Université de Paris VI.

Jacod, J., Protter, P., 1998. Asymptotic error distributions for the Euler method for stochastic differential equations. Annals of Probability 26, 267–307.

Jones, C.S., 2003. The dynamics of stochastic volatility: Evidence from underlying and options markets. Journal of Econometrics 116, 181–224.

LeBaron, B., 2001. Stochastic volatility as a simple generator of financial power laws and long memory. Quantitative Finance 1, 621–631.

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

Meddahi, N., 2001. An eigenfunction approach for volatility modeling. Tech. Rep. Université de Montréal.

Mincer, J., Zarnowitz, V., 1969. The evaluation of economic forecasts. In: Mincer, J. (Ed.), Economic Forecasts and Expectations. National Bureau of Economic Research, New York.

Müller, U.A., Dacorogna, M.M., Davé, R.D., Olsen, R.B., Pictet, O.V., von Weizsäcker, J.E., 1997. Volatilities of different time resolutions — analyzing the dynamics of market components. Journal of Empirical Finance 4, 213–239.

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Oomen, R.C.A., 2006. Properties of realized variance under alternative sampling schemes. Journal of Business and Economic Statistics 24, 219–237.

Pipiras, V., Taqqu, M., 2000. Integration questions related to fractional Brownian motion. Probability Theory and Related Fields 118, 251–291.

Robinson, P.M., 1991. Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. Journal of Econometrics 47, 67–84.

Robinson, P.M., 1995. Log-periodogram regression of time series with long range dependence. Annals of Statistics 23, 1048–1072.

Schwert, G.W., 1987. Effects of model specification on tests for unit roots in macroeconomic data. Journal of Monetary Economics 20, 73–103.

Zhang, L., 2006. Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. Bernoulli 12, 1019–1043.

Zhang, L., Mykland, P.A., Aït-Sahalia, Y., 2005a. Edgeworth expansions for realized volatility and related estimators. Tech. Rep. Princeton University.

Zhang, L., Mykland, P.A., Aït-Sahalia, Y., 2005b. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. Journal of the American Statistical Association 100, 1394–1411.

Zhou, B., 1996. High-frequency data and volatility in foreign-exchange rates. Journal of Business and Economic Statistics 14, 45–52.

Zumbach, G., Corsi, F., Trapletti, A., 2002. Efficient estimation of volatility using high frequency data. Tech. Rep. Olsen & Associates.