Data-based ranking of realised volatility estimators

Andrew J. Patton *

Department of Economics, Duke University, USA
Oxford-Man Institute of Quantitative Finance, University of Oxford, UK

ARTICLE INFO

ABSTRACT

This paper presents new methods for comparing the accuracy of estimators of the quadratic variation of a price process. I provide conditions under which the relative accuracy of competing estimators can be consistently estimated (as $T \to \infty$), and show that forecast evaluation tests may be adapted to the problem of ranking these estimators. The proposed methods avoid making specific assumptions about microstructure noise, and facilitate comparisons of estimators that would be difficult using methods from the extant literature, such as those based on different sampling schemes. An application to high frequency IBM data between 1996 and 2007 illustrates the new methods.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

The past decade has seen an explosion in research on volatility measurement, as distinct from volatility forecasting.[1] This research has focused on constructing non-parametric estimators of price variability over some horizon (for example, one day) using data sampled at a shorter horizons (for example, every 5 minutes or every 30 seconds). These "realised volatility" (RV) estimators or "realised measures" generally aim at measuring the quadratic variation or integrated variance of the log-price process of some asset or collection of assets.

This profusion of research has lead to a need for some practical guidance on which RV estimator to select for a given empirical analysis. In addition to the particular estimator to use, the performance of RV estimators is generally affected by the frequency used to sample the price process (for example, every 5 minutes or every 30 seconds), see Zhou (1996) and Bandi and Russell (2008) for example, and may also be affected by the decision to sample in calendar time or in "tick time" (for example,

every r minutes or every s trades), and the decision to use prices from transactions or from quotes, see Bandi and Russell (2006b), Hansen and Lunde (2006a) and Oomen (2006).

This paper provides new methods for comparing RV estimators, which complement the approaches currently in the literature (discussed further below). Denoting the latent quadratic variation of the process over some interval of time (for example, one day) as θ_t, estimators of this quantity as X_{it}, and a distance measure as L, the primary theoretical contribution of this paper is to provide methods to consistently estimate:

$$E\left[\Delta L\left(\theta_t, \mathbf{X}_t\right)\right] \equiv E\left[L\left(\theta_t, X_{it}\right)\right] - E\left[L\left(\theta_t, X_{jt}\right)\right]. \tag{1}$$

The latent nature of θ_t makes estimating $E\left[\Delta L\left(\theta_t, \mathbf{X}_t\right)\right]$ more difficult than in standard forecasting applications, as we cannot employ the sample mean of the loss differences as an estimator. Further, the fact that the estimators X_{it} usually use data from the same period over which θ_t is measured makes this problem distinct from (and more difficult than) volatility forecasting applications.

With an estimator of $E\left[\Delta L\left(\theta_t, \mathbf{X}_t\right)\right]$ in hand, it is possible to employ one of the many tests from the literature on forecast evaluation and comparison, such as Diebold and Mariano (1995) and West (1996) for pair-wise comparisons, White (2000), Hansen (2005), Hansen et al. (forthcoming) and Romano and Wolf (2005) for comparisons involving a large number of RV estimators, and Giacomini and White (2006) for conditional comparisons of RV estimators. These tests rely on standard large sample asymptotics ($T \to \infty$) rather than continuous-record asymptotics ($m \to \infty$), and thus can be used to compare the "finite m" performance

---

* Corresponding address: Department of Economics, Duke University, 213 Social Sciences Building, Box 90097, Durham, NC 27708-0097, USA. Tel.: +1 919 660 1849.
E-mail address: andrew.patton@duke.edu.

[1] See Andersen and Bollerslev (1998), Andersen et al. (2001a, 2003), Barndorff-Nielsen and Shephard (2002, 2004a), Aït-Sahalia et al. (2005), Zhang et al. (2005), Hansen and Lunde (2006a), Christensen and Podolskij (2007), and Barndorff-Nielsen et al. (2008) amongst many others. Andersen et al. (2006) and Barndorff-Nielsen and Shephard (2007) present recent surveys of this burgeoning field.

of different estimators. I provide conditions under which these tests can be applied to the problem of ranking RV estimators. The proposed methods rely on the existence of a volatility proxy that is unbiased for the latent target variable, θ_t, and satisfies an uncorrelatedness condition, described in detail below. This proxy must be unbiased but it does not need to be very precise; a simple and widely-available proxy is the daily squared return, for example.

Previous research on the selection of estimators of quadratic variation has predominantly focused on finding the sampling frequency that maximises the accuracy of a given estimator. Consider the simplest RV estimator:

$$\mathrm{RV}_t^{(m)} \equiv \sum_{i=1}^{m} \left(p_{\tau_i} - p_{\tau_{i-1}} \right)^2 \tag{2}$$

where p_{τ_i} is the log-price at time τ_i, $\{\tau_0, \tau_1, \ldots, \tau_m\}$ are the times at which the price of the asset is available during period t, and m is the number of intra-period observations used in computing the estimator. In the absence of market microstructure effects, the distribution theory for the simplest RV estimator would suggest sampling prices as often as possible, see Andersen et al. (2001a) for example, as the asymptotic variance of the estimator in this case declines uniformly as $m \to \infty$. In practice, however, the presence of autocorrelation in very high frequency prices leads the standard RV estimator to become severely biased,[2] and several papers have attempted to address this problem.[3] While the methods of these papers differ, they have in common their use of continuous-record asymptotics in their derivations, the use of mean squared error (MSE) as the measure of accuracy, and, importantly, generally quite specific assumptions about the noise process.[4,5]

In contrast to the theoretical studies of the optimal sampling frequency cited above, the data-based methods proposed in this paper allow one to avoid taking a stand on some important properties of the price process. In particular, the proposed approach allows for microstructure noise that may be correlated with the efficient price process and/or heteroskedastic, *cf.* Hansen and Lunde (2006a), Kalnina and Linton (2008), and Bandi et al.

(2007). Further, this approach avoids the need to estimate quantities such as the integrated quarticity and the variance of the noise process, which often enter formulas for the optimal sampling frequency, see Andersen et al. (2011b) and Bandi and Russell (2008) for example, and which can be difficult to estimate in practice. This approach does, however, require some assumptions about the time series properties of the variables under analysis (e.g., stationarity of certain functions of variables), which are not required in most of the existing literature, and so the proposed tests complement, rather than substitute, existing methods; they provide an alternate approach to addressing the same important problem.

The data-based methods proposed in this paper also allow for comparisons of estimators of quadratic variation that would be difficult using existing theoretical methods in the literature. For example, theoretical comparisons of estimators using quote prices versus trade prices require assumptions about the behaviour of market participants: the arrival rate of trades, the placing and removing of limit and market orders, etc., and theoretical comparisons may be sensitive to these assumptions. Likewise, theoretical comparisons of tick-time and calendar-time sampling requires assumptions on the arrival rate of trades. Finally, the methods of this paper make it possible to compare estimators based on quite different assumptions about the price process, such as the "alternation" estimator of Large (2011) which is based on the assumption that the price process moves in steps of at most one tick, versus, for example, the multi-scale estimator of Zhang (2006), which is based on a quite different set of assumptions.

The methods for comparing the accuracy of RV estimators proposed in this paper complement recent work comparing the accuracy of forecasts based on these estimators, see Andersen et al. (2003), Aït-Sahalia and Mancini (2008), and Ghysels and Sinko (2011), among others. If the forecasting model in which the estimator will be used is known by the econometrician, then rankings of RV estimators by their forecast performance are likely of primary interest. However, if the forecasting model is not known by the econometrician, or if the end-use of the RV estimator is unknown more generally (e.g., it may be used in pricing derivatives, risk management, portfolio decisions etc.) then a measure of its estimation accuracy may be of interest as a general gauge of its quality as a proxy for the true, latent, volatility. Of course, the methods proposed in this paper may be combined with measures of forecast accuracy to obtain overall rankings of RV estimators.

The main empirical contribution of this paper comes from a study of the problem of estimating the daily quadratic variation of IBM equity prices, using high frequency data over the period January 1996 to June 2007. I consider simple realised variance estimators based on either quote or trade prices, sampled in either calendar time or in tick time, for many different sampling frequencies. Also studied are four more sophisticated estimators of QV. I find that Romano and Wolf (2005) tests clearly reject the squared daily return in favour of an RV estimator using higher frequency data, and corresponding tests also indicate that there are significant gains to moving beyond the rule-of-thumb of using 5-min calendar-time RV: estimators based on data sampled at between 15 s and 2 min are significantly more accurate than 5-min RV. I also find that some of the more sophisticated estimators of QV proposed in the literature significantly outperform 5-min RV, particularly in the latest sub-sample. In general, I find that using tick-time sampling leads to better estimators than using calendar-time sampling, particularly when trade arrivals are very irregularly-spaced. I also find that quote prices are significantly less accurate that trade prices in the early part of the sample, but this difference disappears in the most recent sub-sample.

---

[2] Early research in this area, see Zhou (1996) and Andersen et al. (2000), employed "volatility signature plots" to show graphically that at very high frequencies, features such as bid–ask bounce and stale prices can lead to large biases in simple RV estimators. More sophisticated estimators, such as the two-scale estimator of Zhang et al. (2006) and the realised kernel estimator of Barndorff-Nielsen et al. (2008) provide consistent estimates of quadratic variation, under some conditions, by taking these autocorrelations into account in the construction of the estimator.

[3] Assuming *i.i.d.* noise and intra-daily homoskedasticity, Zhou (1996) derives the MSE-optimal sampling frequency (or, equivalently, optimal choice of m) for the *RVAC1* estimator, which adjusts the standard RV estimator to account for autocovariances up to order 1; Aït-Sahalia et al. (2005) derive the MSE-optimal choice of m for the standard RV estimator under a variety of cases (*i.i.d.* noise, serially correlated noise, and noise correlated with the efficient price); Andersen et al. (2011b) derive the MSE-optimal choice of m for the *RV ACq* estimator, the realised kernel estimator of Barndorff-Nielsen et al. (2008) and the two-scale estimator of Zhang et al. (2005), under the assumption of *i.i.d.* noise; Hansen and Lunde (2006a) derive the MSE-optimal choice of m for *RV ACq* estimators assuming *i.i.d.* noise; Bandi and Russell (2006a, 2011) derive the optimal choice of the m for standard RV, and the optimal ratio of q/m for *RV ACq* estimators using m intra-daily observations, under the assumption of *i.i.d.* noise; Bandi et al. (2007) consider the optimal choice of m when the noise process is conditionally mean zero but potentially heteroskedastic; and Barndorff-Nielsen et al. (2008) examine the optimal sampling frequency and number of lags to use with a variety of realised kernel estimators, under the assumption of *i.i.d.* noise.

[4] Gatheral and Oomen (2010) provide an alternative analysis of the problem of choosing an RV estimator via a detailed simulation study.

[5] It should be noted that several of these papers derive the asymptotic distribution of their estimators, as $m \to \infty$, under weaker assumptions on the noise than are required to derive optimal sampling frequencies.

The remainder of the paper is structured as follows. Section 2 presents the main theoretical results of this paper, Section 3 presents a simulation study of the proposed new methods, and Section 4 presents an application using high frequency quote and trade data on IBM over the period January 1996–June 2007. Section 5 concludes, and all proofs are collected in the Appendix.

2. Data-based ranking of RV estimators

2.1. Notation and background

The target variable, generally quadratic variation (QV) or integrated variance[6] (IV), is denoted θ_t. I assume that θ_t is $\mathcal{F}_t$-measurable, where $\mathcal{F}_t$ is the information set generated by the complete path of the log-price process. For the remainder of the paper I assume that θ_t is a scalar; I discuss the extension to vector (or matrix) target variables in the conclusion. The estimators of θ_t are denoted X_{it}, $i = 1, 2, \ldots, k$. Often these will be the same estimator applied to data sampled at different frequencies, for example 1-min returns versus 30-min returns, though they could also be RV estimators based on different functional forms, different sampling schemes, etc.

In order to rank the competing estimators we need some measure of distance from the estimator, X_{it}, to the target variable, θ_t. Two popular (pseudo-)distance measures in the volatility literature are MSE and QLIKE:

MSE $\quad L(\theta, X) = (\theta - X)^2$ (3)

QLIKE $\quad L(\theta, X) = \frac{\theta}{X} - \log\left(\frac{\theta}{X}\right) - 1.$ (4)

The definition of QLIKE above has been normalised to yield a distance of zero when $\theta = X$. The methods below apply to rankings of RV estimators using the general class of "robust" pseudo-distance measures proposed in Patton (2011), which nests MSE and QLIKE as special cases:

$L(\theta, X) = \tilde{C}(X) - \tilde{C}(\theta) + C(X)(\theta - X)$ (5)

with C being some function that is decreasing and twice-differentiable function on the supports of both arguments of this function, and where $\tilde{C}$ is the anti-derivative of C. In this class each pseudo-distance measure L is completely determined by the choice of C. MSE and QLIKE are obtained (up to location and scale constants) when $C(z) = -z$ and $C(z) = 1/z$ respectively.

For the remainder of the paper I will use the following notation to describe the $(k - 1$ vector of) differences in the distances from the target variable to a collection of RV estimators:

$\Delta L(\cdot, \mathbf{X}_t) \equiv [L(\cdot, X_{1t}) - L(\cdot, X_{2t}), \ldots, L(\cdot, X_{1t}) - L(\cdot, X_{kt})]'$ (6)

where

$\mathbf{X}_t \equiv [X_{1t}, \ldots, X_{kt}]'.$

Throughout, variables denoted with a '*' below are the bootstrap samples of the original variables obtained from the stationary bootstrap, P is the original probability measure, and P^* is the probability measure induced by the bootstrap conditional on the original data.

2.2. Ranking volatility forecasts versus ranking RV estimators

Ranking volatility forecasts, as opposed to estimators, has received a lot of attention in the econometrics literature, see Poon and Granger (2003) and Hansen and Lunde (2005) for two recent and comprehensive studies, and this is the natural starting point for considering the ranking of realised volatility

estimators. Hansen and Lunde (2006b) and Patton (2011) show that rankings of volatility *forecasts* using a "robust" loss function and a conditionally unbiased volatility proxy are asymptotically equivalent to rankings using the true latent target variable—this is stated formally in part (a) of the proposition below. Part (b) shows that this result does not hold for rankings of volatility *estimators*, due to a critical change in the time at which they are observable. In a slight abuse of notation, the proposition below uses θ_t to denote conditional variance in part (a) and quadratic variation in part (b).

Proposition 1. *Let θ_t be the latent scalar quantity of interest, let $\mathcal{F}_t$ be the information set generated by the complete path of the log-price process until time t, and let $\tilde{\mathcal{F}}_t \subset \mathcal{F}_t$ be the information set available to the econometrician at time t. Let (X_{1t}, X_{2t}) be two estimators of θ_t, and let $\tilde{\theta}_t$ be the proxy for θ_t.*

(a) *[Volatility forecasting] If $\theta_t \in \mathcal{F}_{t-1}$, $(X_{1t}, X_{2t}) \in \tilde{\mathcal{F}}_{t-1}$, $\tilde{\theta}_t \in \tilde{\mathcal{F}}_t$ and $E\left[\tilde{\theta}_t | \mathcal{F}_{t-1}\right] = \theta_t$, and if L is a member of the class of distance measures in Eq. (5), then*

$E[L(\theta_t, X_{1t})] \lesseqgtr E[L(\theta_t, X_{2t})]$

$\Leftrightarrow E\left[L\left(\tilde{\theta}_t, X_{1t}\right)\right] \lesseqgtr E\left[L\left(\tilde{\theta}_t, X_{2t}\right)\right].$

(b) *[Volatility estimation] If $\theta_t \in \mathcal{F}_t$, $(X_{1t}, X_{2t}) \in \tilde{\mathcal{F}}_t$, $\tilde{\theta}_t \in \tilde{\mathcal{F}}_t$ and $E\left[\tilde{\theta}_t | \mathcal{F}_{t-1}, \theta_t\right] = \theta_t$, and if L is a member of the class of distance measures in Eq. (5), then*

$E[L(\theta_t, X_{1t})] \lesseqgtr E[L(\theta_t, X_{2t})]$

$\not\Leftrightarrow E\left[L\left(\tilde{\theta}_t, X_{1t}\right)\right] \lesseqgtr E\left[L\left(\tilde{\theta}_t, X_{2t}\right)\right].$

All proofs are presented in the Appendix. The reason the equivalence holds in part (a) but fails in part (b) is that estimation error in (X_{1t}, X_{2t}) will generally be correlated with the error in $\tilde{\theta}_t$ in the latter case. This means that the ranking of RV estimators needs to be treated differently to the ranking of volatility forecasts, and it is to this that we now turn.

2.3. Ranking RV estimators

In this section we obtain methods to consistently estimate the difference in average accuracy of competing estimators of quadratic variation, $E[\Delta L(\theta_t, \mathbf{X}_t)]$, by exploiting some well-known empirical properties of the behaviour of θ_t and by making use of a (function of a) proxy for θ_t, denoted $\tilde{\theta}_t$. This proxy may itself be a RV estimator, of course, and it may be a noisy estimate of the latent target variable, but it must be conditionally unbiased.

Assumption P1. $\tilde{\theta}_t = \theta_t + \nu_t$, with $E[\nu_t | \mathcal{F}_{t-1}, \theta_t] = 0$, a.s.

For many assets the squared daily return can reasonably be assumed to be conditionally unbiased: the mean return is generally negligible at the daily frequency, and the impact of market microstructure effects is often also negligible in daily returns. It should be noted, however, that the presence of jumps in the data generating process will affect the inference obtained using the daily squared return as a proxy: in this case we can compare the estimators in terms of their ability to estimate quadratic variation, which is the integrated variance plus the sum of squared jumps in many cases, see Barndorff-Nielsen and Shephard (2007) for example, but not in terms of their ability to estimate the integrated variance alone. If an estimator of the integrated variance that is conditionally unbiased, for finite m, in the presence of jumps is available, however, then the methods presented below apply directly.

Assumption P2. $Y_t = \sum_{i=1}^{J} \omega_i \tilde{\theta}_{t+i}$, where $1 \leq J < \infty$, $\omega_i \geq 0 \; \forall i$ and $\sum_{i=1}^{J} \omega_i = 1$.

---

[6] Broadly stated, the quadratic variation of a process coincides with its integrated variance if the process does not exhibit jumps, see Barndorff-Nielsen and Shephard (2007) for example.

In the propositions below I consider using a convex combination of leads of $\tilde{\theta}_t$, as in Assumption P2, the simplest special case of which is just a one-period lead (and so $Y_t = \tilde{\theta}_{t+1}$). Using leads of the proxy is important for breaking the correlated measurement errors problem, which makes it possible to overcome the problems identified in Proposition 1. Y_t is thus interpretable as an instrument for $\tilde{\theta}_t$. Our focus on differences in average accuracy makes this a *non-linear* instrumental variables problem, and like other such problems it is not sufficient to simply assume that $\mathrm{Corr}\left[Y_t, \tilde{\theta}_t\right] \neq 0$; some more structure is required. I obtain results in this application by considering two alternative approximations of the conditional mean of θ_t.

Numerous papers on the conditional variance (see Bollerslev et al., 1994; Engle and Patton, 2001; Andersen et al., 2006, for example), or integrated variance (see Andersen et al., 2004, 2007) have reported that these quantities are very persistent, close to being (heteroskedastic) random walks. The popular RiskMetrics model, for example, is based on a unit root assumption for the conditional variance, and in recent work Hansen and Lunde (2010) find that the null of a unit root is rejected for almost none of the Dow Jones 30 stocks. Wright (1999), in contrast, provides thorough evidence against the presence of a unit root in daily conditional variance for several assets. Other authors have studied the persistence of volatility via long memory models, see Ding et al. (1993), Aït-Sahalia and Mancini (2008), Corsi (2009), and Maasoumi and McAleer (2008), for example. As an initial approximation to the observed persistence in volatility, consider the following assumption[7]:

Assumption T1. $\theta_t = \theta_{t-1} + \eta_t$, with $E[\eta_t|\mathcal{F}_{t-1}] = 0$, a.s. and $\theta_t > 0$ a.s.

The approximation in Assumption T1 is likely to be poor in applications where the price process is subject to jumps that contribute substantially to the total QV. Previous authors have found that the jump component of daily QV is much less persistent than the IV component, see Andersen et al. (2007, 2011a,b) for example, and in such cases the sum of these components (ie, the QV) may not be well approximated by Assumption T1.

2.3.1. Unconditional rankings of RV estimators

This section presents results that allow the ranking of RV estimators based on unconditional average accuracy, according to some distance measure *L*. Importantly, the methods presented below allow for the comparison of multiple estimators simultaneously, via the tests of White (2000) and Romano and Wolf (2005) for example.

Proposition 2. (a) *Let Assumptions P1, P2 and T1 hold, and let the pseudo-distance measure L belong to the class in Eq. (5). Then*

$$E[\Delta L(\theta_t, \mathbf{X}_t)] = E[\Delta L(Y_t, \mathbf{X}_t)]$$

for any vector of RV estimators, $\mathbf{X}_t$, and any L such that these expectations exist.
(b) *If we further assume A1 and A2 in the Appendix, then:*

$$\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}\Delta L(Y_t, \mathbf{X}_t) - E[\Delta L(\theta_t, \mathbf{X}_t)]\right)$$
$$\rightarrow^d N(0, \Omega_1), \quad as\ T \rightarrow \infty$$

where Ω_1 is given in the proof.
(c) *If B1 in the Appendix also holds then the stationary bootstrap may also be employed, as:*

$$\sup_z \left| P^*\left[\left\|\frac{1}{T}\sum_{t=1}^{T}\Delta L\left(Y_t^*, \mathbf{X}_t^*\right) - \frac{1}{T}\sum_{t=1}^{T}\Delta L(Y_t, \mathbf{X}_t)\right\| \leq z\right]\right.$$
$$\left. - P\left[\left\|\frac{1}{T}\sum_{t=1}^{T}\Delta L(Y_t, \mathbf{X}_t) - E[\Delta L(\theta_t, \mathbf{X}_t)]\right\| \leq z\right]\right|$$
$$\rightarrow^p 0, \quad as\ T \rightarrow \infty.$$

Part (a) of the above proposition shows that it is possible to obtain an unbiased estimate of the difference in the average distance from the latent target variable, θ_t, using a suitably-chosen volatility proxy, under certain conditions. This opens the possibility to use existing methods from the forecast evaluation literature to help us choose between RV estimators.[8] Parts (b) and (c) of the proposition uses the existing forecast evaluation literature to obtain moment and mixing conditions under which we obtain an asymptotic normal distribution for estimates of the differences in average distance. The conditions in part (b) are sufficient to justify the use of Diebold and Mariano (1995) and West (1996)-style tests for pair-wise comparisons of RV estimator accuracy. Part (c) justifies the use of the bootstrap 'reality check' test of White (2000), the 'model confidence set' of Hansen and Lunde (2010), the SPA test of Hansen (2005), and the stepwise multiple testing method of Romano and Wolf (2005), which are based on the stationary bootstrap of Politis and Romano (1994).

The methods proposed above are complements rather than substitutes for existing methods: the assumptions required for the above result are mostly non-overlapping with the conditions usually required for existing comparison methods. For example, the above proposition does not require any assumptions about the underlying price process (subject to the moment and mixing conditions being satisfied), the microstructure noise process, the trade or quote arrival processes, or the arrivals of limit versus market orders. This means that tests based on the above proposition allow for comparisons of RV estimators that would be difficult using existing methods in the literature. However, unlike most existing tests, the above proposition relies on a long time series of data rather than a continuous sample of prices (i.e., $T \rightarrow \infty$ rather than $m \rightarrow \infty$), on mixing and moment conditions, and on the applicability of the random walk approximation for the target variable. In Section 3 below I show that these assumptions are reasonable in three realistic simulation designs.

In the next proposition I substitute Assumption T1 with one which allows the latent target variable, θ_t, to follow a stationary AR(p) process. The work of Meddahi (2003) and Barndorff-Nielsen and Shephard (2002) shows that integrated variance follows an ARMA(p, q) model for a wide variety of stochastic volatility models for the instantaneous volatility, motivating this generalisation of the result based on a random walk approximation in Proposition 2. Whilst allowing for a general ARMA model is possible, I focus on the AR case both for the ease with which this case can be handled, and the fact that it has been found to perform approximately as well as the theoretically optimal ARMA model in realistic scenarios, see Andersen et al. (2004).

Assumption T2. $\theta_t = \phi_0 + \sum_{i=1}^{p} \phi_i \theta_{t-i} + \eta_t$, with $E[\eta_t|\mathcal{F}_{t-1}] = 0$ a.s., $\theta_t > 0$ a.s., $\phi_1 \neq 0$, the matrix Ψ defined in Eq. (36) is invertible, and $\boldsymbol{\phi} \equiv \left[\phi_1, \ldots, \phi_p\right]'$ is such that θ_t is covariance stationary.

When the order of the autoregression is greater than one, I also require Assumption R1, below. This assumption is plausible for most RV estimators in the literature, as they are generally based

---

[7] Strict positivity of θ_t for this random walk process can be ensured if innovation is a strictly positive random variable with variance proportional to θ_{t-1}, for example $\eta_t = \theta_{t-1}(Z_t - 1)$ where $Z_t \sim i.i.d.\ \log N\left(-\sigma_Z^2/2, \sigma_Z^2\right)$. Many other specifications are possible.

[8] Note that if the accuracy of RV estimators varies over time, then this approach allows us to make comparisons only about the *average* accuracy over the sample period. If variables thought to be correlated with the accuracy of a given estimator are known, then the conditional rankings in the next section may instead be used.

on data from a single day, although Barndorff-Nielsen et al. (2004) and Owens and Steigerwald (2007) are two exceptions.

Assumption R1. X_t is independent of ν_{t-j} for all $j > 0$.

Proposition 3. *Let Assumptions P1, P2 and T2 hold, let the pseudo-distance measure L belong to the class in Eq. (5), and let R1 hold if $p > 1$. Further, define* $\mathbf{Q}_0 \equiv \left[\phi_0, \mathbf{0}_p'\right]'$, $Q_1 \equiv \begin{bmatrix} \mathbf{0}_p' & 0 \\ I_p & 0 \end{bmatrix}$, $P \equiv \begin{bmatrix} 1 & -\phi' \\ \mathbf{0}_p & I_p \end{bmatrix}$ *where $\mathbf{0}_p$ is a $p \times 1$ vector of zeros. Then:*

(a) $E\left[\Delta L\left(\theta_t, \mathbf{X}_t\right)\right] = E\left[\Delta L\left(Y_t, \mathbf{X}_t\right)\right] - \boldsymbol{\beta}$

where

$$
\begin{aligned}
\boldsymbol{\beta} = {} & E\left[\Delta C\left(\mathbf{X}_t\right)\right] \sum_{j=1}^{J} \omega_j g_0^{(j)} / g_1^{(j)} \\
& + \sum_{j=1}^{J} \omega_j \left(1 - 1/g_1^{(j)}\right) E\left[\Delta C\left(\mathbf{X}_t\right) \tilde{\theta}_{t+j}\right] \\
& + \sum_{j=1}^{J} \omega_j \sum_{i=2}^{p} \left(g_i^{(j)} / g_1^{(j)}\right) E\left[\Delta C\left(\mathbf{X}_t\right) \tilde{\theta}_{t+1-i}\right]
\end{aligned}
$$

for any vector of RV estimators, $\mathbf{X}_t$, and any L such that these expectations exist. The variable $g_0^{(j)}$ is defined as the first element of the vector $\left(I - \left(P^{-1}Q_1\right)^j\right)\left(I - P^{-1}Q_1\right)^{-1} P^{-1}\mathbf{Q}_0$, and $g_i^{(j)}$ is defined as $(1, i)$ element of the matrix $\left(P^{-1}Q_1\right)^j$.

(b) *If we further assume A1 and A2 in the Appendix hold for the series $\mathbf{B}_t$, defined in Eq. (35), then:*

$$
\begin{aligned}
& \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}\Delta L\left(Y_t, \mathbf{X}_t\right) - \hat{\boldsymbol{\beta}}_T - E\left[\Delta L\left(\theta_t, \mathbf{X}_t\right)\right]\right) \\
& \quad \to^d N\left(0, \Omega_2\right), \quad \text{as } T \to \infty
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_T \equiv {} & \left(\frac{1}{T}\sum_{t=1}^{T}\Delta C\left(\mathbf{X}_t\right)\right)\left(\sum_{j=1}^{J}\omega_j \hat{g}_0^{(j)}/\hat{g}_1^{(j)}\right) \\
& + \sum_{j=1}^{J}\omega_j\left(1 - 1/\hat{g}_1^{(j)}\right)\frac{1}{T-j}\sum_{t=1}^{T-j}\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+j} \\
& + \sum_{j=1}^{J}\omega_j\sum_{i=2}^{p}\hat{g}_i^{(j)}/\hat{g}_1^{(j)}\frac{1}{T+1-i}\sum_{t=i}^{T}\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1-i}
\end{aligned}
$$

where $\hat{g}_i^{(j)}$, $i = 0, 1, \ldots, p$; $j = 1, 2, \ldots, J$ are estimators of $g_i^{(j)}$ described in the proof.

(c) *If B1 in the Appendix also holds then the stationary bootstrap may also be employed, as:*

$$
\begin{aligned}
& \sup_z \left| P^*\left[\left\|\frac{1}{T}\sum_{t=1}^{T}\Delta L\left(Y_t^*, \mathbf{X}_t^*\right) - \hat{\boldsymbol{\beta}}_T^* \right.\right.\right. \\
& \qquad \left.\left. - \frac{1}{T}\sum_{t=1}^{T}\Delta L\left(Y_t, \mathbf{X}_t\right) + \hat{\boldsymbol{\beta}}_T \right\| \le z\right] \\
& \qquad - P\left[\left\|\frac{1}{T}\sum_{t=1}^{T}\Delta L\left(Y_t, \mathbf{X}_t\right) - \hat{\boldsymbol{\beta}}_T - E\left[\Delta L\left(\theta_t, \mathbf{X}_t\right)\right]\right\| \le z\right] \right| \\
& \quad \to^p 0, \quad \text{as } T \to \infty.
\end{aligned}
$$

Proposition 3 relaxes the assumption of a random walk, at the cost of introducing a bias term to the expected loss computed using the proxy. This bias term, however, can be consistently estimated under the assumption that the target variable follows

a stationary AR(p) process. The cost of the added flexibility in allowing for a general AR(p) process for the target variable is the added estimation error induced by having to estimate the AR(p) parameters, and having to estimate additional terms of the form $E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+j}\right]$. This estimation error will lead to reduced power to distinguish between competing RV estimators than would otherwise be the case.

2.3.2. Conditional rankings of RV estimators

In this section we extend the above results to consider expected differences in distance *conditional* on some information set, thus allowing the use of Giacomini and White (2006)-type tests of equal conditional RV estimator accuracy. The null hypothesis in a GW-type test is:

$$
H_0^* : E\left[\Delta L\left(\theta_t, \mathbf{X}_t\right) | \mathcal{G}_{t-1}\right] = 0 \quad \text{a.s. } t = 1, 2, \ldots . \tag{7}
$$

For pair-wise comparisons of forecasts (or RV estimators, in our case), $\Delta L\left(\theta_t, \mathbf{X}_t\right)$ is a scalar and the above null is usually tested by looking at simple regressions of the form:

$$
\Delta L\left(\theta_t, \mathbf{X}_t\right) = \boldsymbol{\alpha}'\mathbf{Z}_{t-1} + e_t \tag{8}
$$

where $\mathbf{Z}_{t-1} \in \mathcal{G}_{t-1}$ is some $q \times 1$ vector of variables thought to be useful for predicting future differences in estimator accuracy, and testing:

$$
H_0 : \boldsymbol{\alpha} = 0 \quad \text{versus} \quad H_a : \boldsymbol{\alpha} \ne 0. \tag{9}
$$

The following proposition provides conditions under which a feasible form of the above regression:

$$
\Delta L\left(Y_t, \mathbf{X}_t\right) = \tilde{\boldsymbol{\alpha}}'\mathbf{Z}_{t-1} + \tilde{e}_t \tag{10}
$$

provides consistent estimates of the parameter $\boldsymbol{\alpha}$ in the infeasible regression.

Proposition 4. (a) *Let Assumptions P1, P2 and T1 hold, and let the pseudo-distance measure L belong to the class in Eq. (5). If $\mathcal{G}_{t-1} \subset \mathcal{F}_t$, then*

$$
\begin{aligned}
& E\left[\Delta L\left(\theta_t, \mathbf{X}_t\right) | \mathcal{G}_{t-1}\right] = E\left[\Delta L\left(Y_t, \mathbf{X}_t\right) | \mathcal{G}_{t-1}\right] \\
& \quad a.s., \ t = 1, 2, \ldots
\end{aligned}
$$

for any vector of RV estimators, $\mathbf{X}_t$, and any L such that these expectations exist.

(b) *Assume $\Delta L\left(\theta_t, \mathbf{X}_t\right)$ is a scalar and denote the OLS estimator of $\tilde{\boldsymbol{\alpha}}$ in Eq. (10) as $\hat{\boldsymbol{\alpha}}_T$. Then if we further assume A3 and A4 in the Appendix:*

$$
\hat{D}_T^{-1/2}\sqrt{T}\left(\hat{\boldsymbol{\alpha}}_T - \boldsymbol{\alpha}\right) \to^d N\left(0, I\right)
$$

where

$$
\hat{D}_T \equiv \hat{M}_T^{-1}\hat{\Omega}_T\hat{M}_T^{-1}, \qquad \hat{M}_T \equiv \frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_{t-1}\mathbf{Z}_{t-1}',
$$

$$
\Omega_T \equiv V\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{Z}_{t-1}e_t\right]
$$

and with $\hat{\Omega}_T$ some symmetric and positive semi-definite estimator such that $\hat{\Omega}_T - \Omega_T \to^p 0$.

Part (a) of the above proposition shows that the corresponding part of Proposition 2 can be generalised to allow for a conditioning set $\mathcal{G}_{t-1} \subset \mathcal{F}_t$ without any additional assumptions. Part (b) shows that the OLS estimator of the feasible GW regression in Eq. (10) is centred on the true parameter in the infeasible regression in Eq. (8), thus enabling GW-type tests. The variance of the OLS estimator will generally be inflated relative to the variance of

the infeasible regression, but nevertheless the variance can be estimated using standard methods.

The above proposition can also be extended to allow the latent target variable, θ_t, to follow a stationary AR(p) process. The proposition below shows that the AR approximation can be accommodated by using an adjusted dependent variable in the GW-type regression. That is, the infeasible regression is again:

$$\Delta L(\theta_t, \mathbf{X}_t) = \boldsymbol{\alpha}' \mathbf{Z}_{t-p} + e_t \tag{11}$$

while the adjusted regression becomes:

$$\widetilde{\Delta L(\theta_t, \mathbf{X}_t)} = \tilde{\boldsymbol{\alpha}}' \mathbf{Z}_{t-p} + \tilde{e}_t. \tag{12}$$

Note that the variable $\mathbf{Z}_t$ must be lagged by (at least) the order of the autoregression, so for an AR(p) the right-hand side of the GW-type regression would contain $\mathbf{Z}_{t-p}$. Under the random walk approximation the adjusted dependent variable is simply $\widetilde{\Delta L(\theta_t, \mathbf{X}_t)} = \Delta L(Y_t, \mathbf{X}_t)$, while under the AR($p$) approximation it will contain terms related to the parameters of the AR(p) model. For example, specialising the proposition below to an AR(1) with $J = 1$ (so that $Y_t = \tilde{\theta}_{t+1}$) we have:

$$\widetilde{\Delta L(\theta_t, \mathbf{X}_t)} = \Delta L\left(\tilde{\theta}_{t+1}, \mathbf{X}_t\right) - \frac{\phi_0}{\phi_1} \Delta C(\mathbf{X}_t)$$
$$+ \frac{1 - \phi_1}{\phi_1} \Delta C(\mathbf{X}_t) \tilde{\theta}_{t+1}. \tag{13}$$

This adjusted dependent variable is constructed such that $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha}$, and thus estimating Eq. (12) by OLS yields a consistent estimator of the unknown true parameter $\boldsymbol{\alpha}$. (Note that if $\phi_0 = 0$ and $\phi_1 = 1$, which corresponds to the random walk case, the adjustment term drops out and we obtain the same result as in Proposition 4.) Of course, the parameters of the AR(p) process must be estimated, leading to a feasible adjusted regression:

$$\widehat{\Delta L(\theta_t, \mathbf{X}_t)} = \widehat{\tilde{\boldsymbol{\alpha}}}' \mathbf{Z}_{t-p} + \widehat{\tilde{e}}_t \tag{14}$$

where

$$\widehat{\Delta L(\theta_t, \mathbf{X}_t)} = \Delta L\left(\tilde{\theta}_{t+1}, \mathbf{X}_t\right) - \frac{\hat{\phi}_{0,T}}{\hat{\phi}_{1,T}} \Delta C(\mathbf{X}_t)$$
$$+ \frac{1 - \hat{\phi}_{1,T}}{\hat{\phi}_{1,T}} \Delta C(\mathbf{X}_t) \tilde{\theta}_{t+1} \tag{15}$$

in the AR(1) and $J = 1$ case. The dependent variable in the feasible adjusted regression depends on estimated AR(p) parameters, and so standard OLS inference cannot be used.

The proposition below considers the more general AR(p) case, with a proxy that may depend on a convex combination of leads of $\tilde{\theta}_t$, and shows how to account for the fact that the adjustment term involves estimated parameters. A strengthening of Assumption R1 is needed for the test of conditional accuracy if the order of the autoregressive approximation is greater than one.

Assumption R1′. X_t is conditionally independent of ν_{t-j} given $\mathcal{F}_{t-j-1}$, for all $j > 0$.

Proposition 5. *Let Assumptions* P1, P2 *and* T2 *hold, let the pseudo-distance measure L belong to the class in Eq.* (5), *and let* R1′ *hold if* $p > 1$. *Let* $\mathbf{Q}_0$, Q_1, P *and* $g_i^{(j)}$ *be defined as in Proposition* 3. *Finally, assume that* $\Delta L(\theta_t, X_t)$ *is a scalar, and define:*

$$\widetilde{\Delta L(\theta_t, \mathbf{X}_t)} \equiv \Delta L(Y_t, \mathbf{X}_t) + \lambda_0 \Delta C(\mathbf{X}_t) + \lambda_1 \Delta C(\mathbf{X}_t) \tilde{\theta}_{t+1}$$
$$+ \sum_{i=2}^{p} \lambda_i \Delta C(\mathbf{X}_t) \tilde{\theta}_{t+1-i}$$

where

$$\lambda_1 = \frac{1}{\phi_1} - \sum_{j=1}^{J} \omega_j \frac{g_1^{(j)}}{\phi_1}, \quad and$$

$$\lambda_i = -\frac{\phi_i}{\phi_1} - \sum_{j=1}^{J} \omega_j \left(g_i^{(j)} - g_1^{(j)} \frac{\phi_i}{\phi_1}\right), \quad i = 0, 2, 3, \ldots, p$$

and

$$\widehat{\Delta L(\theta_t, \mathbf{X}_t)} \equiv \Delta L(Y_t, \mathbf{X}_t) + \hat{\lambda}_{0,T} \Delta C(\mathbf{X}_t)$$
$$+ \hat{\lambda}_{1,T} \Delta C(\mathbf{X}_t) \tilde{\theta}_{t+1} + \sum_{i=2}^{p} \hat{\lambda}_{i,T} \Delta C(\mathbf{X}_t) \tilde{\theta}_{t+1-i}$$

where $\hat{\lambda}_{i,T}$, $i = 0, 1, \ldots, p$ *are the values of* λ_i *based on estimated values for* ϕ_j *and* $g_i^{(j)}$. *Then:*

(a) $E\left[\widehat{\Delta L(\theta_t, \mathbf{X}_t)} \mathbf{Z}_{t-p}\right] = E\left[\Delta L(\theta_t, \mathbf{X}_t) \mathbf{Z}_{t-p}\right]$

for any $\mathbf{Z}_{t-p} \in \mathcal{F}_{t-p}$.

(b) *Denote the OLS parameter estimate of* $\widetilde{\boldsymbol{\alpha}}$ *in Eq.* (14) *as* $\hat{\boldsymbol{\alpha}}_T$. *If we further assume* A1 *and* A2 *in the Appendix hold for the series* $\mathbf{D}_t$, *defined in Eq.* (37), *then:*

$$\sqrt{T}\left(\hat{\boldsymbol{\alpha}}_T - \boldsymbol{\alpha}\right) \to^d N(0, \Omega_3) \quad as\ T \to \infty.$$

(c) *If* B1 *in the Appendix also holds then the stationary bootstrap may also be employed, as:*

$$\sup_z \left|P^*\left[\|\hat{\boldsymbol{\alpha}}_T^* - \hat{\boldsymbol{\alpha}}_T\| \le z\right] - P\left[\|\hat{\boldsymbol{\alpha}}_T - \boldsymbol{\alpha}\| \le z\right]\right| \to^p 0,$$

as $T \to \infty$.

As in Proposition 3, the AR assumption introduces additional terms to be estimated in order to consistently estimate $E\left[\Delta L(\theta_t, X_t) \cdot \mathbf{Z}_{t-p}\right]$. The above proposition shows that these terms are estimable, though the additional estimation error will of course reduce the power of this test. It is worth noting that Proposition 3 can be obtained as a special case of the above proposition by simply setting $\mathbf{Z}_{t-p}$ equal to one.

3. Simulation study

To examine the finite-sample performance of the results in the previous section, I present the results of a small simulation study. I use three different stochastic volatility models, each with the same parameters as in Gonçalves and Meddahi (2009). The first model is a GARCH diffusion:

$$d \log P^\dagger(t) = 0.0314 d(t) + \nu(t)$$
$$\times \left(-0.576 dW_1(t) + \sqrt{1 - 0.576^2} dW_2(t)\right) \tag{16}$$

$$d\nu^2(t) = 0.035\left(0.636 - \nu^2(t)\right) d(t) + 0.144 \nu^2(t) dW_1(t).$$

The second model is a log-normal diffusion, using the same process for the log-price as above, but a different process for the volatility:

$$d \log \nu^2(t) = -0.0136\left(0.8382 + \log \nu^2(t)\right) d(t)$$
$$+ 0.1148 dW_1(t). \tag{17}$$

The third volatility model is a two-factor diffusion, which takes the following form:

$$d \log P^\dagger(t) = 0.030 d(t) + \nu(t)\left(-0.30 dW_1(t) - 0.30 dW_2(t)\right.$$
$$\left. + \sqrt{1 - 2 \times (0.30^2)} dW_3(t)\right)$$

$$\nu^2(t) = \exp\left\{-1.2 + 0.04 \nu_1^2(t) + 1.5 \nu_2^2(t)\right\} \tag{18}$$

$$d\nu_1^2(t) = -0.00137 \nu_1^2(t) dt + dW_1(t)$$

$$d\nu_2^2(t) = -1.386 \nu_2^2(t) dt + \left(1 + 0.25 \nu_2^2(t)\right) dW_2(t).$$

The two-factor diffusion is characterized by one highly persistent component and one less persistent component, which yields volatility dynamics that are quite distinct from the other two processes. We include all three processes in this study to gain a better understanding of the finite sample properties of the proposed tests in a variety of empirical situations.

In simulating from these processes I use a simple Euler discretization scheme, with the step size calibrated to one-tenth of one second (i.e., with 234,000 steps per simulated trade day, which assumed to be 6.5 h in length). I consider sample sizes of $T = 500$ and $T = 2500$ trade days.

To gain some insight into the impact of microstructure effects, I also consider a simple *i.i.d.* error term for the observed log-price:

$$\log P\left(t_j\right) = \log P^\dagger\left(t_j\right) + \xi\left(t_j\right) \tag{19}$$

$$\xi\left(t_j\right) \sim i.i.d. \ N\left(0, \sigma_\xi^2\right).$$

Following Aït-Sahalia et al. (2005) and Huang and Tauchen (2005), I set σ_ξ^2 to be such that the proportion of the variance of the 5-min return (5/390 of a trade day) that is attributable to microstructure noise is 20%:

$$\frac{2\sigma_\xi^2}{V\left[r_t\right]\frac{5}{390} + 2\sigma_\xi^2} = 0.20 \tag{20}$$

where r_t is the open-to-close return on day t. The expression above is from Aït-Sahalia et al. (2005), while the proportion of 20% is around the middle value considered in the simulation study of Huang and Tauchen (2005).

The processes to be simulated above exhibit a leverage effect and are contaminated with noise, and so existing results on the ARMA processes for QV implied by various continuous-time stochastic volatility models, see Barndorff-Nielsen and Shephard (2002) and Meddahi (2003), cannot be directly applied. This allows us to study how the proposed tests perform in realistic cases where both the random walk and AR(p) models are merely approximations to the true process for daily QV; neither is correctly specified.

The finite-sample size and power properties of the proposed methods are investigated via the following experiment. For simplicity I focus on pair-wise comparisons of RV estimators, each implemented using the 1000 draws from the stationary bootstrap of Politis and Romano (1994), thus making this a 'reality check'-type test from White (2000). I set the each RV estimator equal to the true QV plus some noise:

$$X_{it} = \text{QV}_t + \zeta_{it}, \quad i = 1, 2 \tag{21}$$

$$\zeta_{1t} = \omega v_t^{30\,\min} + (1 - \omega)\sigma_u U_{1t} \tag{22}$$

$$\zeta_{2t} = \omega v_t^{30\,\min} + (1 - \omega)\sigma_u U_{2t} + \sqrt{\sigma_{\zeta 2}^2 - \sigma_{\zeta 1}^2}\, U_{3t} \tag{23}$$

$$[U_{1t}, U_{2t}, U_{3t}]' \sim i.i.d. \ N\left(0, I\right)$$

where

$$v_t^{30\,\min} \equiv \text{RV}_t^{30\,\min} - IV_t.$$

The above structure allows the measurement error on each of the RV estimators to be correlated with the proxy measurement error, consistent with what is faced in practice. As a benchmark, I use the measurement errors on $\text{RV}^{30\,\min}$ to generate this correlation, and I set the correlation to be $\rho = \text{Corr}\left[v_t^{30\,\min}, \zeta_{1t}\right] = 0.5$, by setting the parameters $\left(\omega, \sigma_u^2\right)$ using Eqs. (24) and (25). The equations below also allow me to vary the variance of the errors associated with the RV estimators, $\sigma_{\zeta 1}^2$ and $\sigma_{\zeta 2}^2$. In the study of the size of the tests I set $\sigma_{\zeta 1}^2/V\left[\text{QV}_t\right] = \sigma_{\zeta 2}^2/V\left[\text{QV}_t\right] = 0.1$ (and so the variable U_{3t} drops out of Eq. (23)) which is approximately equal

to $V\left[v_t^{30\,\min}\right]/V\left[\text{QV}_t\right]$ in this simulation. To study the power, I fix $\sigma_{\zeta 1}^2/V\left[\text{QV}_t\right] = 0.1$, and let $\sigma_{\zeta 2}^2/V\left[\text{QV}_t\right] = 0.15, 0.2, 0.5, 1$.

$$\omega = \frac{\rho\sigma_{\zeta 1}}{\sigma_v} \tag{24}$$

$$\sigma_u^2 = \frac{\sigma_v^2\sigma_{\zeta 1}^2\left(1 - \rho^2\right)}{\left(\sigma_v - \rho\sigma_{\zeta 1}\right)^2} \tag{25}$$

where

$$V\left[v_t^{30\,\min}\right] \equiv \sigma_v^2.$$

I consider seven unconditional comparison tests in total. The first test is the infeasible test that would be conducted if the true QV were observable. The power of this test represents an upper bound on what one can expect from the feasible tests. I consider feasible tests under both the random walk approximation (using Proposition 2) and an AR(1) approximation (using Proposition 3). I also consider three different volatility proxies: daily squared returns, 30-min RV and the true QV. The latter case is considered to examine the limiting case of a proxy with no error being put through these tests. The rejection frequencies under each scenario are presented in Table 1, using the QLIKE pseudo-distance measure from Eq. (4). The corresponding results for the MSE distance measure are similar and available upon request.

The first row of each panel of Table 1 corresponds to the case when the null hypothesis is satisfied, and thus we expect these figures to be close to 0.05, the nominal size of the tests. For both sample sizes and across all three diffusion models we see that the finite-sample size is reasonable, with rejection frequencies close to 0.05. Most tests appear to be under-sized, meaning that they are conservative tests of the null. Only in the cases of a short time series and the use of the AR approximation does over-rejection of the null occur. The results for the power of this test are as expected: the power of the new tests are worse than would be obtained if the true QV were observable; power is greater when using a longer time series of data; power is worse when a noisier instrument is used (true QV versus 30-min RV versus daily squared returns); and the power of the test based on the AR(1) approximation is worse than that based on the random walk approximation. The AR(1) approximation has little power when the volatility proxy is very noisy and T is small: in that case it appears that the estimation of the AR parameters overwhelms any information about the relative accuracy of the two RV estimators. The power curves for the GARCH and Log diffusions are similar, while the power of the test under the two-factor diffusion is generally lower, a finding consistent with other papers using this model, see Huang and Tauchen (2005).

Next I consider a simulation study of the Giacomini and White (2006)-style conditional comparisons of RV estimators. I use the following design:

$$X_{1t} = \text{QV}_t + \zeta_{1t} \tag{26}$$

$$X_{2t} = \text{QV}_t - \lambda \text{QV}_{t-1} + \zeta_{2t} \tag{27}$$

$$\zeta_{it} = \omega v_t^{30\,\min} + (1 - \omega)\sigma_u U_{it}, \quad i = 1, 2$$

$$[U_{1t}, U_{2t}]' \sim i.i.d. \ N\left(0, I\right).$$

As in the simulation for tests of unconditional accuracy, I choose ω and σ_u^2 such that $\sigma_{\zeta 1}^2/V\left[\text{QV}_t\right] = \sigma_{\zeta 2}^2/V\left[\text{QV}_t\right] = 0.1$ and $\text{Corr}\left[v_t^{30\,\min}, \zeta_{1t}\right] = \text{Corr}\left[v_t^{30\,\min}, \zeta_{2t}\right] = 0.5$. In the study of finite-sample size, I set $\lambda = 0$. To study power, I consider introducing some time-varying bias to the second RV estimator, by letting the parameter $\lambda = 0.1, 0.2, 0.4, 0.8$, and then estimate regressions of the form:

$$L\left(\tilde\theta_{t+1}, X_{1t}\right) - L\left(\tilde\theta_{t+1}, X_{2t}\right) = \alpha_0 + \alpha_1 \log \frac{1}{10}\sum_{j=1}^{10}\tilde\theta_{t-j} + e_t \tag{28}$$

Table 1
Finite-sample size and power of unconditional accuracy tests.

	QV*		QV				RV-30 min				RV-daily			
			RW		AR		RW		AR		RW		AR	
T	500	2500	500	2500	500	2500	500	2500	500	2500	500	2500	500	2500
γ	GARCH diffusion													
0.10	0.02	0.02	0.01	0.01	0.01	0.01	0.02	0.02	0.03	0.06	0.03	0.03	0.10	0.02
0.15	0.30	0.53	0.24	0.49	0.20	0.44	0.19	0.43	0.20	0.51	0.11	0.23	0.09	0.09
0.20	0.50	0.79	0.49	0.78	0.43	0.76	0.41	0.75	0.35	0.77	0.20	0.50	0.09	0.18
0.50	0.75	0.99	0.77	0.99	0.75	0.99	0.74	0.99	0.60	0.95	0.53	0.89	0.20	0.36
1.00	0.84	0.99	0.85	1.00	0.84	0.99	0.83	1.00	0.65	0.97	0.69	0.95	0.24	0.41
	Log diffusion													
0.10	0.02	0.03	0.02	0.01	0.01	0.00	0.02	0.01	0.04	0.04	0.03	0.02	0.09	0.01
0.15	0.14	0.32	0.14	0.22	0.14	0.22	0.13	0.20	0.13	0.23	0.08	0.12	0.09	0.03
0.20	0.28	0.62	0.29	0.53	0.25	0.52	0.25	0.49	0.23	0.44	0.16	0.30	0.10	0.06
0.50	0.59	0.96	0.60	0.96	0.60	0.97	0.58	0.94	0.38	0.81	0.39	0.75	0.16	0.12
1.00	0.68	0.98	0.73	0.98	0.74	0.98	0.71	0.98	0.47	0.86	0.54	0.87	0.20	0.20
	Two factor diffusion													
0.10	0.04	0.05	0.04	0.03	0.16	0.05	0.03	0.03	0.15	0.06	0.04	0.04	0.17	0.12
0.15	0.05	0.13	0.07	0.09	0.22	0.12	0.07	0.08	0.21	0.12	0.07	0.06	0.20	0.18
0.20	0.10	0.21	0.07	0.11	0.24	0.17	0.08	0.09	0.23	0.18	0.05	0.06	0.20	0.19
0.50	0.20	0.53	0.10	0.16	0.33	0.31	0.11	0.16	0.31	0.28	0.07	0.10	0.28	0.29
1.00	0.26	0.70	0.10	0.17	0.38	0.31	0.09	0.16	0.38	0.27	0.07	0.09	0.30	0.31

Notes: This table presents the rejection frequencies for tests of equal accuracy of two competing RV estimators, using the QLIKE pseudo-distance measure. The first two columns correspond to the ideal infeasible case when the true QV is observable. The remaining columns present results when the available volatility proxy has varying degrees of measurement error, under two approximations for the QV (a random walk (RW) and a first-order autoregression (AR)). The three panels correspond to three different specifications of the continuous time diffusion generating the observed returns. All tests are conducted at the 0.05 level, based on 1000 draws from the stationary bootstrap, and each scenario is simulated 1000 times. The null hypothesis of equal average accuracy is satisfied in the first row of each panel, while in the other rows the second RV estimator has greater noise variance ($\gamma \equiv \sigma_{\zeta2}^2 / V[QV_t]$) than the first ($\sigma_{\zeta1}^2 / V[QV_t] = 0.10$).

Table 2
Finite-sample size and power of conditional accuracy tests, slope coefficient t-test.

	QV*		QV				RV-30 min				RV-daily			
			RW		AR		RW		AR		RW		AR	
T	500	2500	500	2500	500	2500	500	2500	500	2500	500	2500	500	2500
λ	GARCH diffusion													
0.00	0.02	0.03	0.03	0.03	0.00	0.01	0.03	0.02	0.01	0.01	0.03	0.03	0.01	0.00
0.10	0.06	0.15	0.08	0.17	0.02	0.07	0.06	0.14	0.01	0.04	0.03	0.05	0.01	0.00
0.20	0.13	0.38	0.20	0.43	0.06	0.31	0.15	0.40	0.03	0.20	0.03	0.08	0.02	0.01
0.40	0.23	0.74	0.32	0.80	0.12	0.69	0.25	0.75	0.06	0.53	0.04	0.12	0.02	0.02
0.80	0.17	0.53	0.14	0.45	0.14	0.54	0.13	0.38	0.13	0.44	0.06	0.10	0.02	0.03
	Log diffusion													
0.00	0.03	0.04	0.03	0.04	0.01	0.00	0.03	0.03	0.01	0.00	0.04	0.04	0.00	0.00
0.10	0.04	0.07	0.05	0.08	0.02	0.02	0.04	0.07	0.01	0.01	0.03	0.04	0.01	0.00
0.20	0.09	0.20	0.10	0.22	0.04	0.08	0.08	0.21	0.02	0.06	0.03	0.08	0.01	0.00
0.40	0.16	0.43	0.19	0.45	0.12	0.27	0.16	0.42	0.06	0.20	0.04	0.12	0.01	0.01
0.80	0.39	0.93	0.37	0.92	0.38	0.93	0.33	0.90	0.33	0.90	0.11	0.41	0.04	0.20
	Two factor diffusion													
0.00	0.04	0.04	0.03	0.03	0.01	0.00	0.04	0.03	0.01	0.01	0.02	0.03	0.01	0.01
0.10	0.04	0.06	0.04	0.05	0.02	0.02	0.04	0.05	0.02	0.01	0.02	0.03	0.01	0.01
0.20	0.03	0.11	0.04	0.12	0.02	0.05	0.04	0.10	0.01	0.04	0.02	0.04	0.01	0.01
0.40	0.05	0.28	0.06	0.25	0.02	0.16	0.05	0.23	0.02	0.11	0.03	0.08	0.01	0.02
0.80	0.06	0.38	0.04	0.30	0.04	0.34	0.03	0.28	0.03	0.23	0.03	0.11	0.01	0.03

Notes: This table presents the rejection frequencies for tests on the slope coefficient in a regression for testing the equal conditional accuracy of two competing RV estimators, using the QLIKE pseudo-distance measure. The first two columns correspond to the ideal infeasible case when the true QV is observable. The remaining columns present results when the available volatility proxy has varying degrees of measurement error, under two approximations for the QV (a random walk (RW) and a first-order autoregression (AR)). The three panels correspond to three different specifications of the continuous time diffusion generating the observed returns. All tests are conducted at the 0.05 level, based on 1000 draws from the stationary bootstrap, and each scenario is simulated 1000 times. The null hypothesis of a zero slope coefficient is satisfied in the first row of each panel, while in the other rows the second RV estimator at time t has time-varying bias equal to $-\lambda \times QV_{t-1}$, and thus the true slope coefficient is non-zero.

where $\tilde{\theta}_t$ is the volatility proxy: daily squared returns, 30-min RV or the true QV. I use Propositions 4 and 5 to consider two tests based on the above regression: a test that the slope coefficient is zero ($\alpha_1 = 0$), or a joint test that both coefficients are zero ($\alpha_0 = \alpha_1 = 0$). Under the random walk approximation, I can estimate these regressions by simple OLS, and I use Newey and West (1987) to obtain the covariance matrix of the estimated parameters. Under the AR(1) approximation I use 1000 draws from the stationary bootstrap. In the interests of space I present

these simulation results only for the QLIKE distance measure, see Tables 2 and 3; results under the MSE distance measure are similar and available on request.

The first row of each panel in Tables 2 and 3 corresponds to the case where the null hypothesis is true. The tests using the random walk approximation are generally close to the nominal size of 0.05, while the tests using the AR(1) approximation appear to be somewhat under-sized, again implying a conservative test of the null. As expected, the power of the tests to detect violations

Table 3
Finite-sample size and power of conditional accuracy tests, joint test.

	QV						RV-30 min				RV-daily			
	QV*		RW		AR		RW		AR		RW		AR	
T	500	2500	500	2500	500	2500	500	2500	500	2500	500	2500	500	2500
λ	GARCH diffusion													
0.00	0.02	0.01	0.02	0.01	0.01	0.00	0.02	0.02	0.03	0.01	0.03	0.03	0.00	0.01
0.10	0.46	0.62	0.44	0.61	0.25	0.40	0.24	0.53	0.12	0.33	0.04	0.15	0.01	0.04
0.20	0.75	0.92	0.75	0.93	0.47	0.63	0.71	0.91	0.43	0.61	0.19	0.67	0.03	0.16
0.40	0.96	1.00	0.95	1.00	0.56	0.89	0.94	1.00	0.56	0.86	0.80	0.98	0.16	0.43
0.80	1.00	1.00	1.00	1.00	0.78	1.00	1.00	1.00	0.76	1.00	0.98	1.00	0.30	0.83
	Log diffusion													
0.00	0.01	0.02	0.02	0.02	0.00	0.00	0.02	0.02	0.01	0.00	0.02	0.02	0.01	0.01
0.10	0.19	0.34	0.18	0.34	0.10	0.15	0.11	0.28	0.05	0.13	0.04	0.10	0.00	0.03
0.20	0.51	0.83	0.52	0.83	0.24	0.60	0.48	0.80	0.22	0.52	0.14	0.44	0.02	0.11
0.40	0.91	1.00	0.90	1.00	0.47	0.93	0.87	1.00	0.47	0.91	0.64	0.94	0.10	0.44
0.80	1.00	1.00	1.00	1.00	0.77	1.00	0.99	1.00	0.75	1.00	0.96	1.00	0.34	0.86
	Two factor diffusion													
0.00	0.02	0.03	0.02	0.03	0.00	0.01	0.02	0.03	0.00	0.01	0.02	0.03	0.00	0.00
0.10	0.05	0.15	0.04	0.12	0.00	0.02	0.04	0.10	0.01	0.02	0.03	0.06	0.00	0.00
0.20	0.18	0.53	0.12	0.38	0.01	0.08	0.10	0.35	0.00	0.05	0.05	0.17	0.00	0.01
0.40	0.39	0.93	0.25	0.75	0.01	0.25	0.23	0.71	0.01	0.18	0.14	0.45	0.00	0.04
0.80	0.56	1.00	0.41	0.95	0.03	0.43	0.38	0.93	0.02	0.32	0.26	0.74	0.02	0.08

Notes: This table presents the rejection frequencies for tests of equal conditional accuracy of two competing RV estimators, using the QLIKE pseudo-distance measure. The first two columns correspond to the ideal infeasible case when the true QV is observable. The remaining columns present results when the available volatility proxy has varying degrees of measurement error, under two approximations for the QV (a random walk (RW) and a first-order autoregression (AR)). The three panels correspond to three different specifications of the continuous time diffusion generating the observed returns. All tests are conducted at the 0.05 level, based on 1000 draws from the stationary bootstrap, and each scenario is simulated 1000 times. The null hypothesis of equal conditional accuracy is satisfied in the first row of each panel, while in the other rows the second RV estimator at time t has time-varying bias equal to $-\lambda \times QV_{t-1}$.
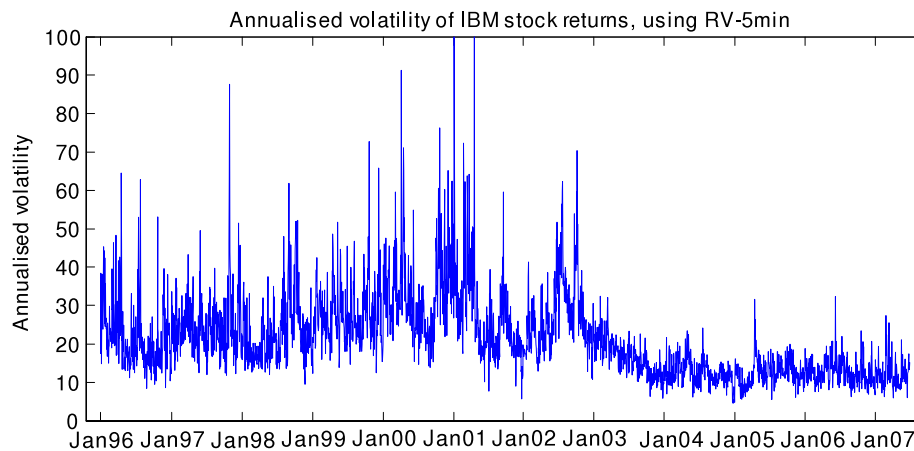

Fig. 1. IBM volatility over the period January 1996–June 2007 (computed using realised volatility based on 5-min calendar-time trade prices), annualised using the formula $\sigma_t = \sqrt{252 \times \mathrm{RV}_t}$.

of the null is lower when a less accurate volatility proxy is employed, higher when a long time series of data is available, and higher using the random walk approximation than using the AR(1) approximation. The results across the three diffusion processes are similar, though again power under the two-factor diffusion is generally lower than under the GARCH or Log diffusion.

4. Estimating the volatility of IBM stock returns

In this section I apply the methods of Section 2 to the problem of estimating the quadratic variation of the open-to-close continuously-compounded return on IBM. I use data on NYSE trade and quote prices from the TAQ database over the period from January 1996 to June 2007, yielding a total of 2893 daily observations.[9] This sample period covers several distinct periods:

the minimum tick size moved from one-eighth of a dollar to one-sixteenth of a dollar on June 24, 1997, and to pennies on January 29, 2001.[10] Further, volatility for this stock (and for the market generally) was high over the early and middle parts of the sample, and very low, by historical standards, in the later years of the sample, see Fig. 1. These changes motivate the use of sub-samples in the empirical analyses below: I break the sample into three periods (1996–1999, 2000–2003 and 2004–2007) to determine whether these changes impact the ranking of the competing realised volatility estimators.

"O", "Z", "B", "T", "L", "G", "W", "J", or "K". Further, the data were cleaned for data problems, following guidelines in Barndorff-Nielsen et al. (2009): trade and quote prices of zero were dropped, as were quotes generating negative spreads or spreads of more than 50 times the median spread for that day. If more than one price was observed with the same time stamp then the median of these prices was used.

9 I use trade and quote prices from the NYSE only, between 9:45 am and 4:00 pm, with a *g127* code of 0 or 40, a *corr* code of 0 or 1, positive *size*, and *cond* not equal to
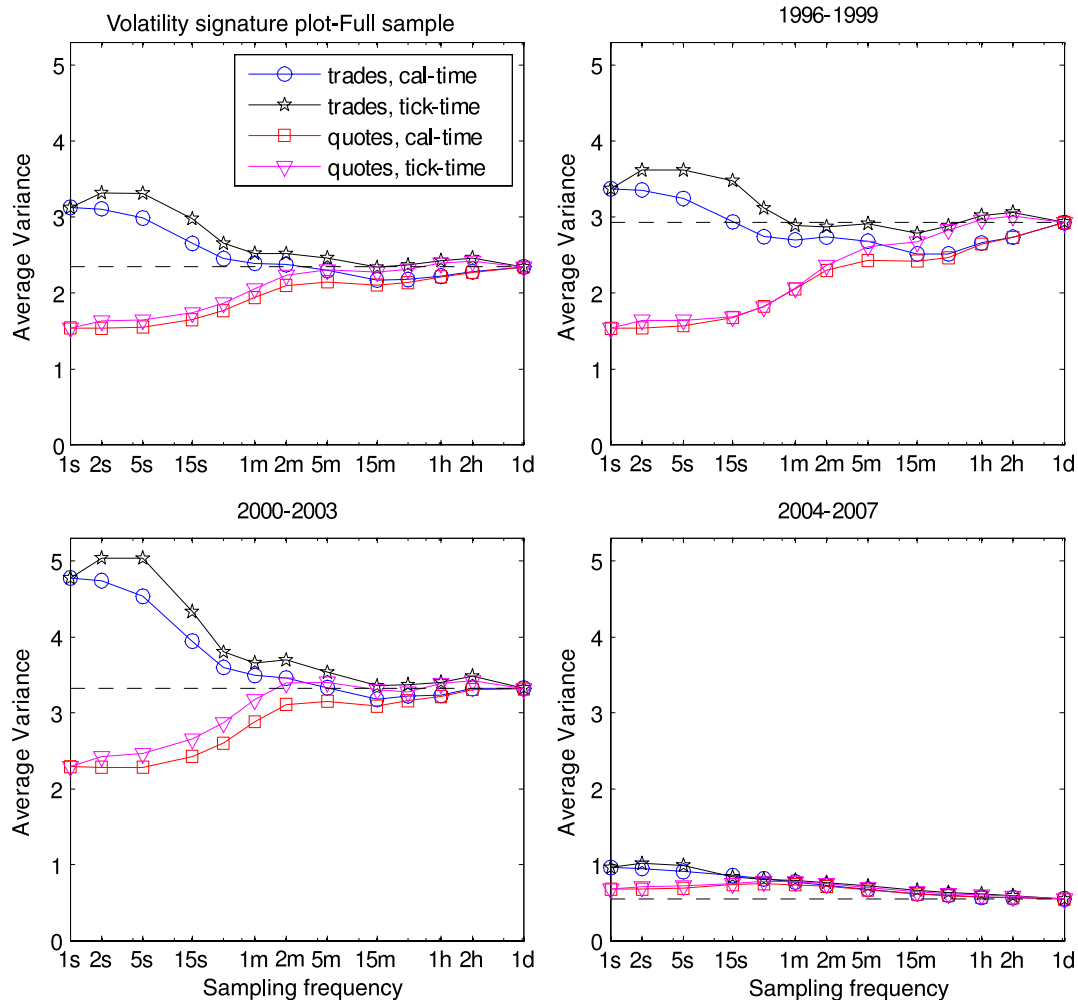
10 Source: New York Stock Exchange web site, http://www.nyse.com/about/history/timeline_chronology_index.html.

Fig. 2. 'Volatility signature plots' for IBM, over the period January 1996–June 2007, using 13 different sampling frequencies (from 1 second to 1 trade day), 2 different price series (trades and quotes) and 2 different sampling schemes (calendar-time and tick-time).

I consider standard realised variance, as presented in Eq. (2), using trade prices and mid-quote prices, and using calendar-time sampling and tick-time sampling, for thirteen different sampling frequencies: 1, 2, 5, 15, 30 seconds, 1, 2, 5, 15, 30 minutes, 1, 2 hours[11] and the open–close return. For tick-time sampling, the sampling frequencies here are *average* times between observations on each day, and the actual sampling frequency of course varies according to the arrival rate of observations. The combination of two price series (trades and mid-quotes), two sampling schemes (calendar-time and tick-time), and 13 sampling frequencies yields 52 possible RV estimators. However, calendar-time and tick-time sampling are equivalent for the two extreme sampling frequencies (1-s sampling and 1-day sampling) which brings the number of RV estimators to 48 in total. In Fig. 2 I present the volatility signature plot for these estimators for the full sample, and for three sub-samples. These plots generally take a common shape: RV computed on trade prices tends to be upward biased for very high sampling frequencies, while RV computed on quote prices tends to be downward biased for very high sampling frequencies, see Hansen and Lunde (2006a) for example. This pattern does not appear in the last sub-sample for this stock.

In Figs. 3 and 4 I present the first empirical contribution of this paper. These figures present estimates of the average distance

between each of the 48 RV estimators and the latent quadratic variation of the IBM price process, relative to the corresponding distance using 5-min calendar-time RV on trade prices,[12] using the QLIKE distance measure presented in Eq. (4).[13] The first figure uses the random walk (RW) approximation for the dynamics in QV, the second uses a first-order AR approximation.[14] I use a one-period lead of 5-min calendar-time RV on trade prices as the volatility proxy to compute the differences in average distances.[15] I

---

[11] I use 62.5 and 125 min sampling rather than 60 and 120 min sampling so that there are an integer number of such periods per trade day. I call these 1-h and 2-h sampling frequencies for simplicity.

[12] The choice of RV estimator to use as the "benchmark" in these plots is purely a normalisation: it has no effect on the ranks of the different estimators.

[13] Patton and Sheppard (2009a) present evidence that the QLIKE pseudo-distance has greater power than the MSE distance measure in a variety of volatility applications. The results of this section under MSE distance are available on request.

[14] The point estimate of the AR coefficient for QV, obtained using the estimator in the proof of Proposition 3, is 0.891, suggesting a strongly persistent volatility process. I estimate the contribution of jumps to QV using the ratio (RV-BV)/RV, where RV is 5-min realised variance and BV is 5-min bipower variation, see Barndorff-Nielsen and Shephard (2004b), which is a jump-robust estimator of IV. This ratio averages 0.07 over this sample period, consistent with Huang and Tauchen (2005) and Tauchen and Zhou (2011), indicating that jumps contribute a small but non-zero amount to the QV of this stock.

[15] Using the assumption that the squared open-to-close return is unbiased for the true quadratic variation, I tested whether 5-min calendar-time RV is also unbiased, and found no evidence against this assumption at the 0.05 level. Using the squared open-to-close return as the volatility proxy did not qualitatively change these results, though as expected the power of the tests was reduced.
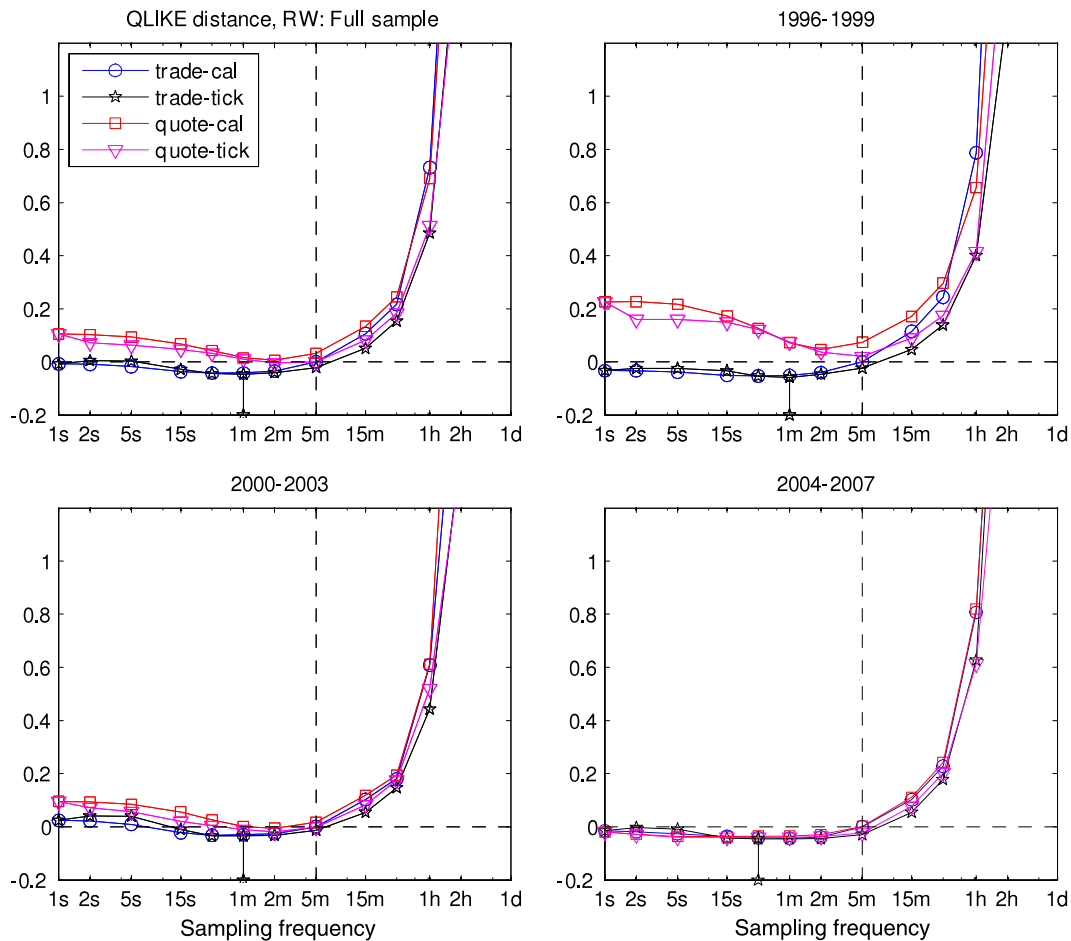
Fig. 3. Differences in average distance, estimated using a random walk approximation, for the 48 competing RV estimators, relative to 5-min calendar-time RV on trade prices. A negative (positive) value indicates that the RV estimator is better (worse) than 5-min calendar-time RV on trade prices. The estimator with the lowest average distance is marked with a vertical line down to the *x*-axis.

present these results for the full sample and for three sub-samples (1996–1999, 2000–2003, 2004–2007).

The conclusion from these pictures is that there are clear gains to using intra-daily data to compute RV, consistent with the voluminous literature to date: the estimated average distances to the true QV for estimators based on returns sampled at 30-min or lower frequencies are clearly greater than those using higher-frequency data (formal tests of this result are presented below). Using the RW approximation, the optimal sampling frequency is either 30 s or 1 min, and the best-performing estimator over the full sample is RV based on trade prices sampled in tick time at 1-min average intervals. The AR approximation gives the same result for the full sample and similar results in the sub-samples.

4.1. Comparing many RV estimators

To formally compare the 48 competing RV estimators, I use the stepwise multiple testing method of Romano and Wolf (2005). This method identifies the estimators that are significantly better, or significantly worse, than a given benchmark estimator, while controlling the family-wise error rate of the complete set of hypothesis tests. That is, for a given benchmark estimator, $X_{t,0}$, it tests:

$$H_0^{(s)} : E\left[L\left(\theta_t, X_{t,0}\right)\right] = E\left[L\left(\theta_t, X_{t,s}\right)\right], \quad \text{for } s = 1, 2, \ldots, 47$$

versus

$$H_1^{(s)} : E\left[L\left(\theta_t, X_{t,0}\right)\right] > E\left[L\left(\theta_t, X_{t,s}\right)\right]$$

or

$$H_2^{(s)} : E\left[L\left(\theta_t, X_{t,0}\right)\right] < E\left[L\left(\theta_t, X_{t,s}\right)\right]$$

and identifies which individual null hypotheses, $H_0^{(s)}$, can be rejected. I use 1000 draws from the stationary bootstrap of Politis and Romano (1994), with an average block size of 20, for each test.

I consider two choices of "benchmark" RV estimators: the squared open-to-close return, which is the most commonly-used volatility estimator in the absence of higher frequency data, and an RV estimator based on 5-min calendar-time trade prices, which is based on a rule-of-thumb from early papers in the RV literature (see Andersen et al., 2001b and Barndorff-Nielsen and Shephard, 2002 for example), which suggests sampling "often but not too often", so as to avoid the adverse impact of microstructure effects.

Table 4 reveals that *every* estimator, except for the squared open-to-close quote-price return, is significantly better than squared open-to-close trade-price return, at the 0.05 level. This is true in the full sample and in all three sub-samples, using both the RW approximation and the AR approximation. This is very strong support for using high frequency data to estimate volatility.

Table 5 provides some evidence that the 5-min RV estimator is significantly beaten by higher-frequency RV estimators. Under the RW approximation, the Romano–Wolf method indicates that RV estimators based on 15-s to 2-min sampling frequencies are significantly better than 5-min RV. Estimators with even higher sampling frequencies are not significantly different, while estimators based on 15-min or lower sampling are found to be
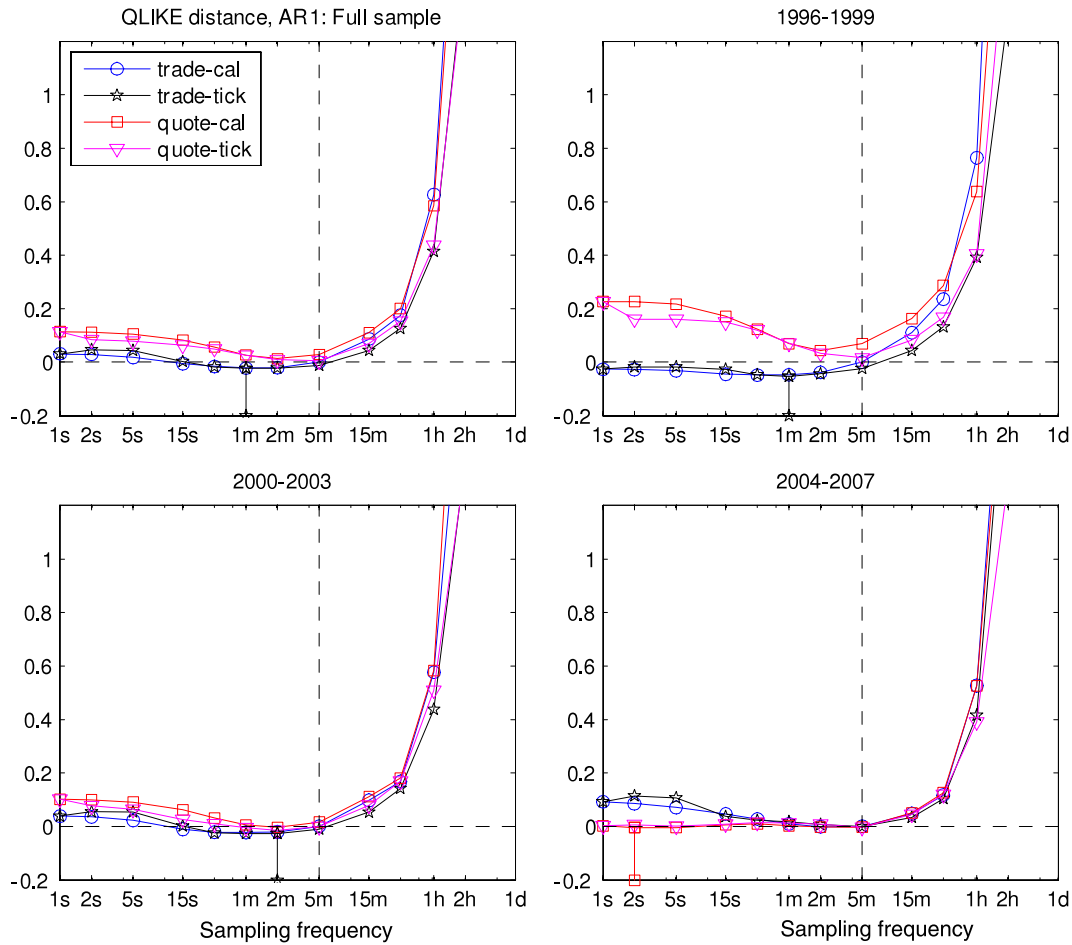
Fig. 4. Differences in average distance, estimated using an AR(1) approximation, for the 48 competing RV estimators, relative to 5-min calendar-time RV on trade prices. A negative (positive) value indicates that the RV estimator is better (worse) than 5-min calendar-time RV on trade prices. The estimator with the lowest average distance is marked with a vertical line down to the *x*-axis.

Table 4
Tests of equal RV accuracy, with squared open-to-close returns as the benchmark.

Sampling frequency	RW approximation				AR approximation			
	Trades		Quotes		Trades		Quotes	
	Calendar	Tick	Calendar	Tick	Calendar	Tick	Calendar	Tick
1 s	✓✓✓✓	–	✓✓✓✓	–	✓✓✓✓	–	✓✓✓✓	–
2 s	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓
5 s	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓
15 s	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓
30 s	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓
1 min	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓
2 min	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓
5 min	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓
15 min	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓
30 min	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓
1 h	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓
2 h	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓
1 day	★	–	~~~~	–	★	–	~~~~	–

Notes: This table presents the results of Romano and Wolf (2005) stepwise testing of the 48 realised volatility estimators considered in the paper (13 frequencies, 2 sampling schemes, 2 price series, less overlaps which are marked with "–"). Two approximations for the dynamics of QV are considered: a random walk (RW) and a first-order autoregression (AR). In this table the benchmark RV estimator is the squared open-to-close trade price return, marked with an ★. Estimators that are significantly better than the benchmark, at the 0.05 level, are marked with ✓, estimators that are significantly worse than the benchmark are marked with ×, and estimators that are not significantly different are marked with ~. The four characters in each element of the above table correspond to the results of the test for the full sample (1996–2007), first sub-sample (1996–1999), second sub-sample (2000–2003) and third sub-sample (2004–2007) respectively.

significantly worse. The results also indicate that trade prices are preferred to quote prices for most of this sample period. Only in the last sub-sample are quote prices at 15-s to 2-min sampling frequencies found to out-perform 5-min RV using trade prices. In the earlier sub-samples quote prices were almost always

worse than trade prices. This result will be explored further in the analysis below. Under the AR approximation very few RV estimators could be distinguished from the 5-min RV estimator using the Romano–Wolf method, suggesting that the gains from moving beyond 5-min sampling are hard to identify in the presence

Table 5
Tests of equal RV accuracy, with the 5-min RV as benchmark.

Sampling frequency	RW approximation				AR approximation			
	Trades		Quotes		Trades		Quotes	
	Calendar	Tick	Calendar	Tick	Calendar	Tick	Calendar	Tick
1 s	~~~~	–	× × × ~	–	~~ × ~	–	~~~~	–
2 s	~~~~	~~ × ~	× × × ~	× × × ~	~~ × ~	~~ × ~	× ~~~	~~~~
5 s	~~~~	~~ × ~	× × × ✓	× × × ✓	~~ × ~	~~ × ~	× ~~~	× ~~~
15 s	✓~~✓	✓ ~ ✓✓	× × ×✓	×× ~✓	~~~~	~~~~	× ~~~	× ~~~
30 s	✓✓ ~✓	✓✓✓✓	× × ×✓	×× ~✓	~~~~	~~~~	× ~~~	× ~~~
1 min	✓✓✓✓	✓✓✓✓	×× ~✓	~ × ~✓	~~~~	~~~~	× ~~~	× ~~~
2 min	✓✓✓✓	✓✓✓✓	~ × ~✓	~ × ~✓	~~~~	~~~~	× ~~~	~~~~
5 min	★	✓~✓✓	× × × ~	~~~~	★	~~~~	× ~~~	~~~~
15 min	× × × ×	× × × ×	× × × ×	× × × ×	~~~~	~~~~	~~~~	~~~~
30 min	× × × ×	× × × ×	× × × ×	× × × ×	~~~~	~~~~	~~~~	~~~~
1 h	× × × ×	× × × ×	× × × ×	× × × ×	~~~~	~~~~	~~~~	~~~~
2 h	× × × ×	× × × ×	× × × ×	× × × ×	~~~~	~~~~	~~~~	~~~~
1 day	× × × ×	–	× × × ×	–	~~~~	–	~~~~	–

Notes: This table presents the results of Romano and Wolf (2005) stepwise testing of the 48 realised volatility estimators considered in the paper (13 frequencies, 2 sampling schemes, 2 price series, less overlaps which are marked with "– "). Two approximations for the dynamics of QV are considered: a random walk (RW) and a first-order autoregression (AR). In this table the benchmark RV estimator is based on 5-min trade prices sampled in calendar time, marked with an ★. Estimators that are significantly better than the benchmark, at the 0.05 level, are marked with ✓, estimators that are significantly worse than the benchmark are marked with ×, and estimators that are not significantly different are marked with ~. The four characters in each element of the above table correspond to the results of the test for the full sample (1996–2007), first sub-sample (1996–1999), second sub-sample (2000–2003) and third sub-sample (2004–2007) respectively.

of additional estimation error from the AR model, consistent with the simulation results in Section 3.

4.2. Comparing more sophisticated estimators of QV

As noted in the Introduction, the past decade has yielded great progress on the estimation of asset price volatility using high frequency data. The realised volatility estimator in Eq. (2) was the first, and remains the simplest, such estimator. In this Section I compare the performance of a selection of more sophisticated estimators of quadratic variation with simple RV estimates.[16] The first two estimators are the two-scale estimator (TSRV) of Zhang et al. (2005) and the multi-scale estimator (MSRV) of Zhang (2006). These estimators use realised variances computed using more than one sampling frequency, which is shown, under certain conditions, to lead to consistency of the estimator in the presence of noise and to efficiency gains. For TSRV I use one tick as the highest frequency and use the optimal "sparse" sampling frequency presented in that paper. For MSRV I again set one tick as the highest frequency and use the formula from that paper for the frequencies of the other estimates and the weights used to combine these estimates. Next, I consider the "realised kernel" (RK) of Barndorff-Nielsen et al. (2008). Following their empirical application to General Electric stock returns, I use their "modified Tukey–Hanning$_2$" kernel and 1-min tick-time sampling, and choose the bandwidth using the approach in Barndorff-Nielsen et al. (2009). Finally, I consider the "realised range-based variance" (RRV) of Christensen and Podolskij (2007) and Martens and van Dijk (2007). I use 5-min blocks, as in Christensen and Podolskij (2007), with 1-min prices within each block. I compare these estimators with RV based on calendar-time trade prices sampled at 1 s, 5 min and 1 day, which gives a total of seven estimators.[17]

I compare each of these estimators against a RV estimator based on 5-min calendar-time trade prices, using a bootstrap version of the Diebold and Mariano (1995) test. I consider both the RW approximation and the AR approximation, drawing on

Propositions 2 and 3 respectively. The results are shown in Table 6. Under the RW approximation all four of the more sophisticated estimators of QV out-perform simple RV$^{5\,\text{min}}$ over the full sample, with the differences being significant for RK and RRV. Under the AR approximation the significance is reduced, and none of the more sophisticated estimators outperform RV$^{5\,\text{min}}$ in the full sample. It is noteworthy that in the latter sub-sample (2004–2007) all four of the more sophisticated estimators significantly beat RV$^{5\,\text{min}}$, under both the RW and AR approximations, perhaps indicating that in latter periods, when turnover and liquidity are higher, there are greater gains to using more sophisticated estimates of QV.

4.3. Conditional comparisons of RV estimators

To investigate the possible sources of the under- or out-performance of certain RV estimators, I next undertake Giacomini and White (2006)-style tests of conditional estimator accuracy. As discussed in Section 2.3.2, the null hypothesis of interest in a Giacomini–White (GW) test is that two competing RV estimators have equal average accuracy conditional on some information set $\mathcal{G}_{t-1}$, that is:

$$H_0^* : E\left[L\left(\theta_t, X_{t,0}\right) | \mathcal{G}_{t-1}\right] - E\left[L\left(\theta_t, X_{t,s}\right) | \mathcal{G}_{t-1}\right] = 0$$
$$\text{a.s. } t = 1, 2, \ldots.$$

One way to implement a test of this null is via a simple regression:

$$L\left(\theta_t, X_{t,0}\right) - L\left(\theta_t, X_{t,s}\right) = \beta_0 + \beta_1 Z_{t-1} + e_t \tag{29}$$

where $Z_{t-1} \in \mathcal{G}_{t-1}$, and then test the necessary conditions:

$$H_0 : \beta_0 = \beta_1 = 0 \tag{30}$$

versus

$$H_a : \beta_i \neq 0 \quad \text{for some } i = 0, 1.$$

4.3.1. High-frequency versus low-frequency RV estimators

I first use the GW test to examine the states where the gains from using high-frequency data are greatest. One obvious conditioning variable is recent volatility: distribution theory for standard RV estimators, see Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2004a) for example, suggests that RV estimators are less accurate during periods of high volatility, and one might expect that the accuracy gains from using high-frequency data are greatest during volatile periods. Using the RW approximation, I estimate the following regression, and obtain the results below, with robust t-statistics presented in parentheses

---

[16] These estimators were computed using Kevin Sheppard's "Oxford Realized" toolbox for Matlab, http://realized.oxford-man.ox.ac.uk/data/code.

[17] Patton and Sheppard (2009b) consider these, and some other, RV estimators in a study of optimal combinations of estimators of QV. A comprehensive study of the performance across several asset classes of the various estimators of QV is being pursued in Patton and Sheppard (2010).

Table 6

Comparing $RV^{5\,min}$ with more sophisticated estimators.

Estimator	RW approximation					AR approximation				
	Avg ΔL	t-statistics on ΔL				Avg ΔL	t-statistics on ΔL			
		99–07	96–99	00–03	04–07		99–07	96–99	00–03	04–07
$RV^{1\,s}$	−0.01	−1.02	−2.74	1.18	−1.73	0.10	2.44	1.12	2.10	−0.65
$RV^{5\,min}$	0.00	★	★	★	★	0.00	★	★	★	★
$RV^{1\,day}$	29.66	9.77	5.81	5.32	8.93	23.03	6.08	4.50	3.97	8.62
TSRV	−0.00	−0.09	3.26	−0.95	−3.22	0.03	1.67	0.32	0.97	−2.39
MSRV	−0.00	−0.29	3.12	−0.90	−3.52	0.02	1.61	−0.03	0.87	−2.58
RK_{TH2}	−0.01	−2.14	0.91	−1.86	−5.66	0.02	1.32	−1.27	0.61	−4.41
RRV	−0.02	−3.41	0.49	−2.75	−6.56	0.01	0.73	−1.66	0.41	−5.09

Notes: This table presents the results of comparisons of the accuracy of seven estimators of QV: realised variance (RV) sampled at 1-s, 5-min and 1-day, two-scales realised variance (TSRV), multi-scale realised variance (MSRV), realised kernel with Tukey–Hanning kernel (RK_{TH2}), and realised range-based variance (RRV). Two approximations for the dynamics of QV are considered: a random walk (RW) and a first-order autoregression (AR). In this table the benchmark RV estimator is based on 5-min trade prices sampled in calendar time, marked with an ★. The full sample differences in average QLIKE accuracy, relative to $RV^{5\,min}$, are reported in the first column of each panel, with negative (positive) values indicating that the estimator is better (worse) than $RV^{5\,min}$. Diebold–Mariano t-statistics tests of the differences in accuracy are presented in the remaining columns, for the full sample (1996–2007), first sub-sample (1996–1999), second sub-sample (2000–2003) and third sub-sample (2004–2007).

below the parameter estimates[18]:

$$L\left(Y_t, RV_t^{(daily)}\right) - L\left(Y_t, RV_t^{(5\,min)}\right) = \underset{(8.42)}{33.67} + e_t \qquad (31)$$

$$L\left(Y_t, RV_t^{(daily)}\right) - L\left(Y_t, RV_t^{(5\,min)}\right) = \underset{(11.10)}{24.94} + \underset{(2.55)}{17.85}\,Z_{t-1} + e_t \quad (32)$$

where

$$Z_{t-1} = \log \frac{1}{10} \sum_{j=1}^{10} Y_{t-j}.$$

The first of the above regression results show that daily squared returns, $RV_t^{(daily)}$, are less accurate on average than RV based on 5-min sampling. The positive and significant coefficient on lagged volatility in the second regression is consistent with RV distribution theory, and indicates that the relative accuracy of daily squared returns deteriorates during high volatility periods. The p-value from a test that both parameters in the second regression are zero is less than 0.001, indicating a strong rejection of the null of equal conditional accuracy.

Using an AR approximation and the bootstrap methods presented in Proposition 5, very similar results are obtained[19]:

$$L\left(Y_t, RV_t^{(daily)}\right) - L\left(Y_t, RV_t^{(5\,min)}\right) = \underset{(6.69)}{33.54} + e_t \qquad (33)$$

$$L\left(Y_t, RV_t^{(daily)}\right) - L\left(Y_t, RV_t^{(5\,min)}\right) = \underset{(5.75)}{19.93} + \underset{(3.45)}{27.76}\,Z_{t-1} + e_t \quad (34)$$

with bootstrap p-values from tests that the parameters in both models are zero less than 0.001 in both cases.

4.3.2. Tick-time versus calendar-time sampling

I next use the GW test of conditional accuracy to compare calendar-time sampling with tick-time sampling. Theoretical comparisons of tick-time and calendar-time sampling requires assumptions on the arrival rate of trades, while the methods presented in this paper allow us to avoid making any specific assumptions about the trade arrival process. For example, in a parametric "pure jump" model of high frequency asset prices,

Oomen (2006) finds that tick-time sampling leads to more accurate RV estimators than calendar-time sampling when trades arrive at irregular intervals. In general, if the trade arrival rate is correlated with the level of volatility, consistent with the work of Easley and O'Hara (1992), Engle (2000) and Manganelli (2005), then using tick-time sampling serves to make the sampled high-frequency returns closer to homoskedastic, which theoretically should improve the accuracy of RV estimation, see Hansen and Lunde (2006a) and Oomen (2006). I use the log volatility of trade durations to measure how irregularly-spaced trade observations are: this volatility will be zero if trades arrive at evenly-spaced intervals, and increases as trades arrive more irregularly.

I estimate a regression of the difference in the accuracy of a calendar-time RV estimator and a tick-time estimator with the same average sampling frequency, on a constant and the lagged log volatility of trade durations, for each of the frequencies considered in the earlier sections,[20] and present the results in Table 7. The first column of Table 7 reports a Diebold and Mariano (1995)-type test of the difference in unconditional average accuracy, across the sampling frequencies, using the RW approximation. This difference is positive and significant for the highest three frequencies (2, 5 and 15 s) and negative and significant for all but one of other frequencies, indicating that tick-time sampling is better than calendar-time sampling (has smaller average distance from the true QV) for all but the very highest frequencies. Further, Table 7 reveals that for all but one frequency the slope coefficient is negative, and 6 out of 11 are significantly negative, indicating that the accuracy of tick-time RV is even better relative to calendar-time RV when trades arrive more irregularly. The results under the AR approximation are very similar to those under the RW approximation, though with slightly reduced significance.

4.3.3. Quote prices versus trade prices

Finally, I examine the difference in accuracy of RV estimators based on trade prices versus quote prices. Theoretical comparisons of RV estimators using quote prices versus trade prices require assumptions about the behaviour of market participants: the arrival rate of trades, the placing and removing of limit and market orders, etc., and theoretical comparisons may be sensitive to these assumptions. The data-based methods of this paper allow us to avoid such assumptions.

As a simple measure of the potential informativeness of quotes versus trades, I consider using the ratio of the number of quotes

---

[18] Tests for zero autocorrelation in the regression residuals and squared regression residuals, up to the tenth lag, yield p-values of less than 0.01 in all cases, motivating the use of a block bootstrap to capture this serial dependence. The R^2 of the second of these regressions is 0.005.

[19] Tests for zero autocorrelation in the regression residuals and squared regression residuals, up to the tenth lag, yield p-values of less than 0.01 in all cases, again motivating the use of a block bootstrap to capture this serial dependence. The R^2 of the second of these regressions is 0.011.

[20] Calendar-time sampling and tick-time sampling are equivalent for the 1-s and 1-day frequencies, and so these are not reported.

Table 7
Tests of equal unconditional and conditional RV accuracy: tick-time versus calendar-time sampling.

Sampling frequency	RW approximation				AR approximation			
	Uncond	Conditional			Uncond	Conditional		
	Avg (t-stat)	Const (t-stat)	Slope (t-stat)	Joint p-val	Avg (t-stat)	Const (t-stat)	Slope (t-stat)	Joint p-val
2 s	0.01* (13.92)	0.08* (3.52)	−0.01* (−2.73)	0.00	0.03* (4.99)	0.27* (2.98)	−0.04* (−2.81)	0.00
5 s	0.02* (12.95)	0.07* (2.79)	−0.01* (−2.01)	0.00	0.04* (4.73)	0.34* (2.67)	−0.05* (−2.49)	0.00
15 s	0.01* (4.03)	−0.15 (−3.88)	0.03 (3.97)	0.00	0.01 (1.20)	−0.27* (−2.83)	0.05* (2.86)	0.00
30 s	−0.00 (−1.35)	−0.02 (−0.48)	0.00 (0.41)	0.08	−0.00 (−0.54)	−0.14 (−1.16)	0.02 (1.13)	0.06
1 min	−0.00* (−2.80)	0.07* (2.29)	−0.01* (−2.39)	0.01	0.01 (1.11)	0.16 (1.83)	−0.03 (−1.78)	0.01
2 min	−0.01* (−3.31)	0.02 (0.57)	−0.00 (−0.71)	0.00	0.01 (1.17)	0.22* (2.02)	−0.04* (−2.03)	0.02
5 min	−0.02* (−6.28)	−0.01 (−0.10)	−0.00 (−0.23)	0.00	0.02 (0.67)	0.49* (2.09)	−0.09* (−2.15)	0.02
15 min	−0.06* (−6.93)	0.17 (1.03)	−0.04 (−1.35)	0.00	−0.02 (−0.78)	0.69* (2.09)	−0.13 (−2.20)	0.03
30 min	−0.06* (−4.02)	0.70* (2.22)	−0.14* (−2.39)	0.00	−0.02 (−0.49)	1.29* (2.29)	−0.24* (−2.40)	0.01
1 h	−0.25* (−3.59)	1.97 (1.63)	−0.40 (−1.78)	0.00	−0.14 (−1.52)	2.99 (1.92)	−0.56 (−2.00)	0.17
2 h	−1.00* (−2.75)	10.66 (1.94)	−2.10* (−2.10)	0.00	−0.86 (−1.79)	9.24 (1.15)	−1.82 (−1.25)	0.10

Notes: This table presents the estimated difference in average distance of tick-time and calendar-time RV estimators, $L\left(Y_t, \mathrm{RV}_t^{\mathrm{tick}(h)}\right) - L\left(Y_t, \mathrm{RV}_t^{\mathrm{cal}(h)}\right)$, either unconditionally, or via a regression on a constant and one-period lag of the log variance of intra-day trade durations, which is a measure of the irregularity of the arrivals of trade observations. A negative slope coefficient indicates that higher volatility of durations leads to an improvement in the accuracy of the tick-time RV estimator relative to a calendar-time RV estimator using the same (average) frequency. Trade prices are used for all RV estimators. The fourth and eighth columns present the p-values from a chi-squared test that both coefficients are equal to zero. Two approximations for the dynamics of QV are considered: a random walk (RW) and a first-order autoregression (AR). Inference under the RW approximation is based on Newey and West (1987) standard errors, while inference under the AR approximation is based on 1000 samples from the stationary bootstrap. All parameter estimates that are significantly different from zero at the 0.05 level are marked with an asterisk.

Table 8
Tests of equal unconditional conditional RV accuracy: quote prices versus trade prices.

Sampling frequency	RW approximation				AR approximation			
	Uncond	Conditional			Uncond	Conditional		
	Avg (t-stat)	Const (t-stat)	Slope (t-stat)	Joint p-val	Avg (t-stat)	Const (t-stat)	Slope (t-stat)	Joint p-val
1 s	0.11* (9.20)	0.33* (10.74)	−0.14* (−8.66)	0.00	0.02 (0.34)	0.44* (5.48)	−0.27* (−3.83)	0.00
2 s	0.11* (9.29)	0.34* (11.22)	−0.14* (−9.38)	0.00	0.02 (0.51)	0.43* (5.97)	−0.26* (−4.37)	0.00
5 s	0.11* (10.03)	0.34* (12.30)	−0.14* (−10.59)	0.00	0.03 (0.97)	0.41* (6.16)	−0.24* (−4.50)	0.00
15 s	0.11* (11.99)	0.30* (13.75)	−0.13* (−11.52)	0.00	0.05* (2.21)	0.34* (6.67)	−0.18* (−5.06)	0.00
30 s	0.08* (12.56)	0.24* (13.70)	−0.10* (−11.43)	0.00	0.05* (2.66)	0.25* (6.52)	−0.13* (−5.53)	0.00
1 min	0.06* (11.90)	0.16* (12.34)	−0.07* (−10.21)	0.00	0.03* (2.28)	0.16* (5.97)	−0.08* (−5.36)	0.00
2 min	0.04* (11.08)	0.11* (10.30)	−0.05* (−8.63)	0.00	0.02* (2.11)	0.10* (4.40)	−0.05* (−4.33)	0.00
5 min	0.03* (8.72)	0.10* (8.80)	−0.05* (−7.85)	0.00	0.02* (3.25)	0.09* (4.39)	−0.04* (−4.08)	0.00
15 min	0.03* (6.33)	0.07* (5.21)	−0.03* (−4.07)	0.00	0.02* (2.28)	0.07* (3.63)	−0.03* (−3.25)	0.00
30 min	0.03* (4.38)	0.07* (3.17)	−0.03* (−2.36)	0.00	0.03* (4.38)	0.07* (3.17)	−0.03* (−2.36)	0.00
1 h	−0.04 (−0.77)	−0.17 (−0.90)	0.08 (0.93)	0.63	−0.05 (−0.76)	−0.13 (−0.68)	0.05 (0.52)	0.67
2 h	−0.19 (−0.74)	−1.06 (−1.47)	0.55 (1.63)	0.25	−0.21 (−0.74)	−1.09 (−1.34)	0.56 (1.44)	0.37
1 day	1.98 (0.41)	10.70 (0.84)	−5.52 (−0.95)	0.62	1.51 (0.25)	10.76 (0.64)	−5.85 (−0.77)	0.66

Notes: This table presents the estimated difference in average distance of quote-price and trade-price RV estimators, $L\left(Y_t, \mathrm{RV}_t^{\mathrm{quote}(h)}\right) - L\left(Y_t, \mathrm{RV}_t^{\mathrm{trade}(h)}\right)$, either unconditionally, or via a regression on a constant and one-period lag of the ratio of the number of quote observations per day to the number of trade observations per day. A negative slope coefficient indicates that an increase in the number of quote observations relative to trades leads to an improvement in the accuracy of the quote-price RV estimator relative to a trade-price RV estimator with the same frequency. Calendar time sampling is used for all estimators. The fourth and eighth columns present the p-values from a chi-squared test that both coefficients are equal to zero. Two approximations for the dynamics of QV are considered: a random walk (RW) and a first-order autoregression (AR). Inference under the RW approximation is based on Newey and West (1987) standard errors, while inference under the AR approximation is based on 1000 samples from the stationary bootstrap. All parameter estimates that are significantly different from zero at the 0.05 level are marked with an asterisk.

per day to the number of trades per day. I regress the difference in the accuracy of a quote-price RV and trade-price RV, with the same calendar-time sampling frequency, on a constant and the lagged ratio of the number of quotes to the number of trades. I do this for each of the frequencies considered in the earlier sections, and present the results in Table 8. The first column of Table 8 reveals that quote-price RV had larger average distance to the true

QV than trade-price RV for all but two sampling frequencies, and for 10 out of 13 this difference is significant at the 0.05 level. However, the results of the test of conditional estimator accuracy reveal that quote-price RV improves relative to trade-price RV as the number of quote observations increases relative to the number of trades: 11 out of 13 slope coefficients are negative, and 10 of these are statistically significant. Results are very similar under the

AR approximation, though with slightly reduced t-statistics. The ratio of quotes per day to trades per day for IBM has increased from around 0.5 in 1996 to around 2.5 in 2007, and may explain the sub-sample results in Table 5: as the relative number of quotes per day has increased, its relative accuracy has also increased. In the early part of the sample, quote-price RV was significantly less accurate than trade-price RV, however that difference vanishes in the last sub-sample, where quote and trade prices, of the same frequency, yield approximately equally accurate RV estimators.[21]

5. Conclusion

This paper considers the problem of ranking competing realised volatility (RV) estimators, motivated by the growing literature on nonparametric estimation of price variability using high-frequency data, see Andersen et al. (2006) and Barndorff-Nielsen and Shephard (2007) for recent surveys. I provide conditions under which the relative average accuracy of competing estimators for the latent target variable can be consistently estimated from available data, using "large T", asymptotics, and show that existing tests from the forecast evaluation literature, such as Diebold and Mariano (1995), West (1996), White (2000), Hansen et al. (forthcoming), Romano and Wolf (2005) and Giacomini and White (2006), may then be applied to the problem of ranking these estimators. The methods proposed in this paper eliminate the need for specific assumptions about the properties of the microstructure noise, and facilitate comparisons of RV estimators that would be difficult using methods from the extant literature.

I apply the proposed methods to high frequency IBM stock price data between 1996 and 2007 in a detailed empirical study. I consider simple RV estimators based on either quote or trade prices, sampled in either calendar-time or in tick-time, for several different sampling frequencies. Romano and Wolf (2005) tests reject the squared daily return and the 5-min calendar-time RV in favour of an RV estimator using data sampled at between 15 s and 5 min. In general, I found that using tick-time sampling leads to more accurate RV estimation than using calendar-time sampling, particularly when trades arrivals are very irregularly-spaced, and RV estimators based on quote prices are significantly less accurate than those based on trade prices in the early part of the sample, but this difference disappears in the most recent sub-sample of the data.

Acknowledgements

---

[21] Using data from the first half of 2007, corresponding to the end of the last sub-sample in this paper, Barndorff-Nielsen et al. (2009) also find that estimators based on quote prices are very similar to those based on trade prices, when kernel-based estimators of the type in Barndorff-Nielsen et al. (2008) are used, or when standard RV estimators are used on slightly-lower frequency data (1–5-min sampling rather than 1-s sampling).

Appendix. Proofs

Additional assumptions used in parts of the proofs below:

Let $\mathbf{A}_t \equiv \left[\Delta L (\theta_t, \mathbf{X}_t)', \Delta C (\mathbf{X}_t)' (Y_t - \theta_t) \right]'$, let $\bar{\mathbf{A}}_T$ denote the sample mean of $\mathbf{A}_t$, and let $A_{i,t}$ denote the ith element of $\mathbf{A}_t$.

Assumption A1. $E \left[\left| A_{i,1} \right|^{6+\varepsilon} \right] < \infty$ for some $\varepsilon > 0$ and for all i.

Assumption A2. $\{ \mathbf{A}_t \}$ is α-mixing of size $-3 (6 + \varepsilon) / \varepsilon$.

Assumption A3. $E [\mathbf{Z}_{t-1} e_t] = 0$ for all t.

Assumption A4(a). $\left\{ \left[\mathbf{Z}'_{t-1}, \tilde{e}_t \right] \right\}$ is α-mixing of size $- (2 + \varepsilon) / \varepsilon$ for some $\varepsilon > 0$.

Assumption A4(b). $E \left[\left| Z_{t-1,i} \tilde{e}_t \right|^{2+\varepsilon} \right] < \infty$ for $i = 1, 2, \ldots, q$ and all t.

Assumption A4(c). $V_T \equiv V \left[T^{-1/2} \sum_{t=1}^{T} \mathbf{Z}_{t-1} \tilde{e}_t \right]$ is uniformly positive definite.

Assumption A4(d). $E \left[\left| Z_{t-1,i} \right|^{2+\varepsilon+2\delta} \right] < \infty$ for some $\delta > 0$ and all $i = 1, 2, \ldots, q$ and all t.

Assumption A4(e). $M_T \equiv E \left[T^{-1} \sum_{t=1}^{T} \mathbf{Z}_{t-1} \mathbf{Z}'_{t-1} \right]$ is uniformly positive definite.

Assumption B1. If p_T is the inverse of the average block length in Politis and Romano's (1994) stationary bootstrap, then $p_T \to 0$ and $T \times p_T \to \infty$.

Proof of Proposition 1. The proof of part (a) is given in Hansen and Lunde (2006b). I repeat part of it here to show where that proof breaks down in part (b). Consider a second-order mean-value expansion of the pseudo-distance measure $L \left(\tilde{\theta}_t, X_{it} \right)$ around (θ_t, X_{it}):

$$
\begin{aligned}
L \left(\tilde{\theta}_t, X_{it} \right) &= L (\theta_t, X_{it}) + \frac{\partial L (\theta_t, X_{it})}{\partial \theta} \left(\tilde{\theta}_t - \theta_t \right) \\
&\quad + \frac{1}{2} \frac{\partial^2 L \left(\ddot{\theta}_t, X_{it} \right)}{\partial \theta^2} \left(\tilde{\theta}_t - \theta_t \right)^2 \\
&= L (\theta_t, X_{it}) + (C (X_{it}) - C (\theta_t)) \left(\tilde{\theta}_t - \theta_t \right) \\
&\quad - \frac{1}{2} C' \left(\ddot{\theta}_t \right) \left(\tilde{\theta}_t - \theta_t \right)^2
\end{aligned}
$$

where $\ddot{\theta}_t = \lambda_t \theta_t + (1 - \lambda_t) \tilde{\theta}_t$ for some $\lambda_t \in [0, 1]$, and using the functional form of L in Eq. (5). The third term in the above equation does not depend on X_{it}, and so will not affect the ranking of (X_{1t}, X_{2t}). In volatility forecasting applications, θ_t is the conditional variance and so $\theta_t \in \mathcal{F}_{t-1}$, and X_{it} is a volatility forecast, and so $X_{it} \in \tilde{\mathcal{F}}_{t-1}$. In that case, this allows

$$
\begin{aligned}
&E \left[(C (X_{it}) - C (\theta_t)) \cdot \left(\tilde{\theta}_t - \theta_t \right) | \mathcal{F}_{t-1} \right] \\
&= (C (X_{it}) - C (\theta_t)) \cdot \left(E \left[\tilde{\theta}_t | \mathcal{F}_{t-1} \right] - \theta_t \right) = 0
\end{aligned}
$$

by the unbiasedness of $\tilde{\theta}_t$ for θ_t conditional on $\mathcal{F}_{t-1}$. Using the law of iterated expectations we obtain $E \left[(C (X_{it}) - C (\theta_t)) \cdot \left(\tilde{\theta}_t - \theta_t \right) \right] = 0$, and thus $E \left[\Delta L \left(\tilde{\theta}_t, \mathbf{X}_t \right) \right] = E [\Delta L (\theta_t, \mathbf{X}_t)]$.

(b) When X_{it} is a realised volatility estimator and θ_t is the integrated variance or quadratic variation we have $\theta_t \in \mathcal{F}_t$ and

$\left(X_{it}, \tilde{\theta}_t \right) \in \tilde{\mathcal{F}}_t$, which means we cannot employ the above reasoning directly. If we could assume that $\text{Corr}\left[C\left(X_{it} \right) - C\left(\theta_t \right), \tilde{\theta}_t - \theta_t | \mathcal{F}_{t-1} \right] = 0 \; \forall i$, in addition to $E\left[\tilde{\theta}_t | \mathcal{F}_t \right] = \theta_t$, then we would have

$$E\left[\left(C\left(X_{it} \right) - C\left(\theta_t \right) \right) \left(\tilde{\theta}_t - \theta_t \right) | \mathcal{F}_{t-1} \right]$$
$$= E\left[C\left(X_{it} \right) - C\left(\theta_t \right) | \mathcal{F}_{t-1} \right] E\left[\tilde{\theta}_t - \theta_t | \mathcal{F}_{t-1} \right]$$
$$= E\left[C\left(X_{it} \right) - C\left(\theta_t \right) | \mathcal{F}_{t-1} \right] E\left[E\left[\tilde{\theta}_t | \mathcal{F}_{t-1}, \theta_t \right] - \theta_t | \mathcal{F}_{t-1} \right]$$
$$= 0.$$

However it is *not* true that $\text{Corr}\left[C\left(X_{it} \right) - C\left(\theta_t \right), \tilde{\theta}_t - \theta_t | \mathcal{F}_{t-1} \right] = 0$ for all empirically relevant combinations of RV estimators and volatility proxies. In fact, if $X_{it} = \tilde{\theta}_t$ and $L = MSE$, a very natural case to consider, then $C(z) = -z$ and $\text{Corr}\left[C\left(X_{it} \right) - C\left(\theta_t \right), \tilde{\theta}_t - \theta_t | \mathcal{F}_{t-1} \right] = \text{Corr}\left[\theta_t - \tilde{\theta}_t, \tilde{\theta}_t - \theta_t | \mathcal{F}_{t-1} \right] = -1$. In general, we should expect $\text{Corr}\left[C\left(X_{it} \right) - C\left(\theta_t \right), \tilde{\theta}_t - \theta_t | \mathcal{F}_{t-1} \right] \neq 0$. This is the correlation between the error in $\tilde{\theta}_t$ and something similar to the "generalised forecast error", see Patton and Timmermann (2010) for example, of X_{it}. If the proxy, $\tilde{\theta}_t$, and the RV estimators, X_{it}, use the same or similar data then their errors will generally be correlated and this zero correlation restriction will not hold, and thus $E\left[\left(C\left(X_{it} \right) - C\left(\theta_t \right) \right) \left(\tilde{\theta}_t - \theta_t \right) \right] \neq 0$, which breaks the equivalence of the ranking obtained using $\tilde{\theta}_t$ with that using θ_t.   □

Proof of Proposition 2. (a) See the proof of part (a) Proposition 4 and set $\mathcal{G}_{t-1}$ to be the trivial information set.
(b) Note that

$$\frac{1}{T}\sum_{t=1}^{T} \Delta L\left(Y_t, \mathbf{X}_t \right) - E\left[\Delta L\left(\theta_t, \mathbf{X}_t \right) \right]$$
$$= \frac{1}{T}\sum_{t=1}^{T} \Delta L\left(\theta_t, \mathbf{X}_t \right) - E\left[\Delta L\left(\theta_t, \mathbf{X}_t \right) \right]$$
$$\quad + \frac{1}{T}\sum_{t=1}^{T} \{ \Delta L\left(Y_t, \mathbf{X}_t \right) - \Delta L\left(\theta_t, \mathbf{X}_t \right) \}$$
$$= \frac{1}{T}\sum_{t=1}^{T} \Delta L\left(\theta_t, \mathbf{X}_t \right) - E\left[\Delta L\left(\theta_t, \mathbf{X}_t \right) \right]$$
$$\quad + \frac{1}{T}\sum_{t=1}^{T} \Delta C\left(\mathbf{X}_t \right)\left(Y_t - \theta_t \right)$$
$$\equiv \bar{\mathbf{A}}_T - E\left[\mathbf{A}_t \right]$$

where

$$\mathbf{A}_t \equiv \left[\Delta L\left(\theta_t, \mathbf{X}_t \right)', \Delta C\left(\mathbf{X}_t \right)'\left(Y_t - \theta_t \right) \right]'$$

since $E\left[\Delta C\left(\mathbf{X}_t \right)\left(Y_t - \theta_t \right) \right] = 0$ from part (a). Under Assumptions A1 and A2, Theorem 3 of Politis and Romano (1994) provides:

$$\sqrt{T}\left(\bar{\mathbf{A}}_T - E\left[\mathbf{A}_t \right] \right) \to^d N\left(0, V_A \right)$$

where V_A is the long-run covariance matrix of $\mathbf{A}_t$. Let ι denote a vector of ones, and note that

$$\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} \Delta L\left(Y_t, \mathbf{X}_t \right) - E\left[\Delta L\left(\theta_t, \mathbf{X}_t \right) \right] \right)$$
$$= \iota' \sqrt{T}\left(\bar{\mathbf{A}}_T - E\left[\mathbf{A}_t \right] \right) \to^d N\left(0, \Omega_1 \right)$$

where $\Omega_1 \equiv \iota' V_A \iota$. It should be noted that Assumptions A1 and A2 can hold despite the random walk Assumption T1, if θ_t and X_{it} obey a some form of cointegration, linked to the distance measure

employed. If MSE is employed, T1, A1 and A2 require that these variables obey standard linear cointegration, with cointegrating vector $[1, -1]$. For other distance measures a form of non-linear cointegration must hold.
(c) Follows directly from Theorem 3 of Politis and Romano (1994), under the additional Assumption B1.   □

Proof of Proposition 3. (a) Using the second-order mean-value expansion of the loss function from the proof of Proposition 4, we obtain $E\left[\Delta L\left(Y_t, \mathbf{X}_t \right) \right] = E\left[\Delta L\left(\theta_t, \mathbf{X}_t \right) \right] + \boldsymbol{\beta}$, where $\boldsymbol{\beta} \equiv E\left[\Delta C\left(\mathbf{X}_t \right)\left(Y_t - \theta_t \right) \right] = \sum_{j=1}^{J} \omega_j E\left[\Delta C\left(\mathbf{X}_t \right)\left(\tilde{\theta}_{t+j} - \theta_t \right) \right] = \sum_{j=1}^{J} \omega_j E\left[\Delta C\left(\mathbf{X}_t \right)\left(\theta_{t+j} - \theta_t \right) \right] = \sum_{j=1}^{J} \omega_j E\left[\Delta C\left(\mathbf{X}_t \right)\left(E_t\left[\theta_{t+j} \right] - \theta_t \right) \right]$ under P1 and P2. Allowing for $J > 1$ requires computing $j \; (> 1)$-step ahead forecasts from an AR(p) process, $E_t\left[\theta_{t+j} \right]$. This is simplified by using the companion form for the AR(p) process governing θ_t:

$$\begin{bmatrix} 1 & -\phi_1 & \cdots & -\phi_p \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \theta_t \\ \theta_{t-1} \\ \vdots \\ \theta_{t-p} \end{bmatrix}$$
$$= \begin{bmatrix} \phi_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \theta_{t-1} \\ \theta_{t-2} \\ \vdots \\ \theta_{t-p-1} \end{bmatrix} + \begin{bmatrix} \nu_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

redefine as

$$P\mathbf{Z}_t = \mathbf{Q}_0 + Q_1\mathbf{Z}_{t-1} + \mathbf{V}_t$$

so

$$\mathbf{Z}_t = P^{-1}\mathbf{Q}_0 + P^{-1}Q_1\mathbf{Z}_{t-1} + P^{-1}\mathbf{V}_t$$

with

$$E\left[\mathbf{Z}_t \right] = \left(I - P^{-1}Q_1 \right)^{-1} P^{-1}\mathbf{Q}_0$$

and so

$$E_t\left[\mathbf{Z}_{t+j} \right] = \left(I - P^{-1}Q_1 \right)^{-1} P^{-1}\mathbf{Q}_0 + \left(P^{-1}Q_1 \right)^j$$
$$\quad \times \left(\mathbf{Z}_t - \left(I - P^{-1}Q_1 \right)^{-1} P^{-1}\mathbf{Q}_0 \right)$$
$$= \left\{ \left(I - \left(P^{-1}Q_1 \right)^j \right) \left(I - P^{-1}Q_1 \right)^{-1} P^{-1}\mathbf{Q}_0 \right\} + \left(P^{-1}Q_1 \right)^j \mathbf{Z}_t$$

and so

$$E_t\left[\theta_{t+j} \right] = g_0^{(j)} + \sum_{i=1}^{p} g_i^{(j)} \theta_{t+1-i}$$

where $g_0^{(j)}$ is the first element of $\left(I - \left(P^{-1}Q_1 \right)^j \right)\left(I - P^{-1}Q_1 \right)^{-1} P^{-1}\mathbf{Q}_0$, and $g_i^{(j)}$ is the $(1, i)$ element of $\left(P^{-1}Q_1 \right)^j$. Next I use this result to obtain:

$$E\left[\Delta C\left(\mathbf{X}_t \right) E_t\left[\theta_{t+j} \right] \right] = g_0^{(j)} E\left[\Delta C\left(\mathbf{X}_t \right) \right] + g_1^{(j)} E\left[\Delta C\left(\mathbf{X}_t \right) \theta_t \right]$$
$$\quad + \sum_{i=2}^{p} g_i^{(j)} E\left[\Delta C\left(\mathbf{X}_t \right) \theta_{t+1-i} \right]$$

so

$$E\left[\Delta C\left(\mathbf{X}_t \right) \theta_t \right] = \frac{1}{g_1^{(j)}} E\left[\Delta C\left(\mathbf{X}_t \right) \tilde{\theta}_{t+j} \right] - \frac{g_0^{(j)}}{g_1^{(j)}} E\left[\Delta C\left(\mathbf{X}_t \right) \right]$$
$$\quad - \sum_{i=2}^{p} \frac{g_i^{(j)}}{g_1^{(j)}} E\left[\Delta C\left(\mathbf{X}_t \right) \theta_{t+1-i} \right]$$

which yields

$$E\left[\Delta C\left(\mathbf{X}_t\right)\left(\tilde{\theta}_{t+j} - \theta_t\right)\right]$$

$$= \left(1 - \frac{1}{g_1^{(j)}}\right) E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+j}\right] + \frac{g_0^{(j)}}{g_1^{(j)}} E\left[\Delta C\left(\mathbf{X}_t\right)\right]$$

$$+ \sum_{i=2}^{p} \frac{g_i^{(j)}}{g_1^{(j)}} E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1-i}\right]$$

since $E\left[\Delta C\left(\mathbf{X}_t\right)\theta_{t+1-i}\right] = E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1-i}\right]$ for $i \geq 2$ under Assumption R1. With this result we can now compute $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} = \sum_{j=1}^{J} \omega_j E\left[\Delta C\left(\mathbf{X}_t\right)\left(E_t\left[\theta_{t+j}\right] - \theta_t\right)\right]$$

$$= \sum_{j=1}^{J} \omega_j \left(1 - \frac{1}{g_1^{(j)}}\right) E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+j}\right]$$

$$+ \sum_{j=1}^{J} \omega_j \frac{g_0^{(j)}}{g_1^{(j)}} E\left[\Delta C\left(\mathbf{X}_t\right)\theta_t\right]$$

$$+ \sum_{j=1}^{J} \omega_j \sum_{i=2}^{p} \frac{g_i^{(j)}}{g_1^{(j)}} E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1-i}\right].$$

Substituting in this expression for $\boldsymbol{\beta}$, we thus have

$$E\left[\Delta L\left(\theta_t, \mathbf{X}_t\right)\right] = E\left[\Delta L\left(Y_t, \mathbf{X}_t\right)\right]$$

$$- \sum_{j=1}^{J} \omega_j \left(1 - \frac{1}{g_1^{(j)}}\right) E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+j}\right]$$

$$- \sum_{j=1}^{J} \omega_j \frac{g_0^{(j)}}{g_1^{(j)}} E\left[\Delta C\left(\mathbf{X}_t\right)\right] - \sum_{j=1}^{J} \omega_j \sum_{i=2}^{p} \frac{g_i^{(j)}}{g_1^{(j)}} E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1-i}\right].$$

(b) This is proved by invoking a multivariate CLT for the sample mean of the loss differentials using the true volatility and all of the elements that enter into the estimated bias term, $\hat{\beta}_T$. This collection of elements is defined as:

$$\mathbf{B}_t \equiv \left[\Delta L\left(\theta_t, \mathbf{X}_t\right)', \Delta C\left(\mathbf{X}_t\right)', \Delta C\left(\mathbf{X}_t\right)'\tilde{\theta}_{t+1}, \ldots, \Delta C\left(\mathbf{X}_t\right)'\tilde{\theta}_{t+J},\right.$$

$$\Delta C\left(\mathbf{X}_t\right)'\tilde{\theta}_{t-1}, \ldots, \Delta C\left(\mathbf{X}_t\right)'\tilde{\theta}_{t-p+1}, \tilde{\theta}_t,$$

$$\left.\tilde{\theta}_t\tilde{\theta}_{t+1}, \ldots, \tilde{\theta}_t\tilde{\theta}_{t+2p}\right]'$$

(35)

and with Assumptions A1 and A2 applied to $\mathbf{B}_t$ we have $\sqrt{T}\left(\bar{\mathbf{B}}_T - E\left[\mathbf{B}_t\right]\right) \to^d N\left(0, V_B\right)$ using Theorem 3 of Politis and Romano (1994).

Note that the last $2p + 1$ elements of $\bar{\mathbf{B}}_T$ are sufficient to obtain estimates of the mean and the first $2p$ autocovariances of θ_t, since $E\left[\tilde{\theta}_t\right] = E\left[\theta_t\right]$ by Assumption P1, and $E\left[\tilde{\theta}_t\tilde{\theta}_{t+j}\right] = E\left[\left(\theta_t + v_t\right)\left(\theta_{t+j} + v_{t+j}\right)\right] = E\left[\theta_t\theta_{t+j}\right]$ by Assumptions P1 and T2. Let $\gamma_j \equiv \text{Cov}\left[\theta_t, \theta_{t-j}\right]$, then by the properties of an AR(p) process we have $\Psi\boldsymbol{\phi} = \boldsymbol{\psi}$, where

$$\Psi \equiv \begin{bmatrix} \gamma_p & \gamma_{p-1} & \cdots & \gamma_1 \\ \gamma_{p+1} & \gamma_p & \cdots & \gamma_2 \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{2p-1} & \gamma_{2p-2} & \cdots & \gamma_p \end{bmatrix},$$

(36)

$$\boldsymbol{\psi} = \left[\gamma_{p+1}, \gamma_{p+2}, \ldots, \gamma_{2p}\right]', \qquad \boldsymbol{\phi} = \left[\phi_1, \ldots, \phi_p\right]'$$

and by Assumption T2 we can obtain $\hat{\boldsymbol{\phi}} = \hat{\Psi}^{-1}\hat{\boldsymbol{\psi}}$, where $\hat{\Psi}$ and $\hat{\boldsymbol{\psi}}$ are the equivalents of Ψ and $\boldsymbol{\psi}$ using sample autocovariances

rather than population autocovariances. From $\hat{\boldsymbol{\phi}}$ we can obtain estimates of P, $\mathbf{Q}_0$, and Q_1 and thus estimates of the parameters $g_i^{(j)}$, for $i = 0, 1, \ldots, p$ and $j = 1, 2, \ldots, J$, from these we can compute the estimated bias term $\hat{\boldsymbol{\beta}}_T$. Given asymptotic normality of $\bar{\mathbf{B}}_T$ and the fact that $\frac{1}{T}\sum_{t=1}^{T}\Delta L\left(Y_t, \mathbf{X}_t\right) - \hat{\boldsymbol{\beta}}_T$ is a smooth function of the elements of $\bar{\mathbf{B}}_T$, we can then apply the delta method, see Lemma 2.5 of Hayashi (2000) for example, to obtain asymptotic normality of $\frac{1}{T}\sum_{t=1}^{T}\Delta L\left(Y_t, \mathbf{X}_t\right) - \hat{\boldsymbol{\beta}}_T$ and obtain its covariance matrix.
(c) Follows directly from Theorem 4 of Politis and Romano (1994), under the additional Assumption B1. □

Proof of Proposition 4. (a) Consider again a second-order mean-value expansion of the pseudo-distance measure $L\left(Y_t, X_{it}\right)$ given in Eq. (5) around $\left(\theta_t, X_{it}\right)$:

$$L\left(Y_t, X_{it}\right) = L\left(\theta_t, X_{it}\right) + \frac{\partial L\left(\theta_t, X_{it}\right)}{\partial \theta}\left(Y_t - \theta_t\right)$$

$$+ \frac{1}{2}\frac{\partial^2 L\left(\ddot{\theta}_t, X_{it}\right)}{\partial \theta^2}\left(Y_t - \theta_t\right)^2$$

$$= L\left(\theta_t, X_{it}\right) + \left(C\left(X_{it}\right) - C\left(\theta_t\right)\right)\left(Y_t - \theta_t\right)$$

$$- \frac{1}{2}C'\left(\ddot{\theta}_t\right)\left(Y_t - \theta_t\right)^2,$$

where $\ddot{\theta}_t = \lambda_t\theta_t + \left(1 - \lambda_t\right)Y_t$ for some $\lambda_t \in \left[0, 1\right]$, and using the functional form of L in Eq. (5). Thus

$$\Delta L\left(Y_t, \mathbf{X}_t\right) = \Delta L\left(\theta_t, \mathbf{X}_t\right) + \Delta C\left(\mathbf{X}_t\right)\left(Y_t - \theta_t\right)$$

where

$$\Delta C\left(\mathbf{X}_t\right) \equiv \begin{bmatrix} C\left(X_{1t}\right) - C\left(X_{2t}\right) \\ \vdots \\ C\left(X_{1t}\right) - C\left(X_{kt}\right) \end{bmatrix}.$$

Next, note:

$$E\left[\Delta C\left(\mathbf{X}_t\right)\left(Y_t - \theta_t\right)|\mathcal{G}_{t-1}\right]$$

$$= E\left[\Delta C\left(\mathbf{X}_t\right)\left(\sum_{i=1}^{J}\omega_i\tilde{\theta}_{t+i} - \theta_t\right)\middle|\mathcal{G}_{t-1}\right]$$

$$= E\left[\Delta C\left(\mathbf{X}_t\right)\left(\sum_{i=1}^{J}\omega_i\sum_{j=1}^{i}\eta_{t+j} + \sum_{i=1}^{J}\omega_i v_{t+i}\right)\middle|\mathcal{G}_{t-1}\right]$$

$$= E\left[\Delta C\left(\mathbf{X}_t\right)\left(\sum_{i=1}^{J}\omega_i\sum_{j=1}^{i}E\left[\eta_{t+j}|\mathcal{F}_t\right]\right.\right.$$

$$\left.\left.+ \sum_{i=1}^{J}\omega_i E\left[v_{t+i}|\mathcal{F}_t\right]\right)\middle|\mathcal{G}_{t-1}\right] = 0$$

by the law of iterated expectations, since $\mathcal{G}_{t-1} \subset \mathcal{F}_t$. This thus yields $E\left[\Delta L\left(\theta_t, \mathbf{X}_t\right)|\mathcal{G}_{t-1}\right] = E\left[\Delta L\left(Y_t, \mathbf{X}_t\right)|\mathcal{G}_{t-1}\right]$ as claimed.
(b) Using Exercise 5.21 of White (2001) for example, we have $\hat{D}_T^{-1/2}\sqrt{T}\left(\hat{\boldsymbol{\alpha}}_T - \tilde{\boldsymbol{\alpha}}\right) \to^d N\left(0, I\right)$, where $\hat{D}_T$ is given in the statement of the proposition.

To show that $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha}$ note that $\Delta L\left(Y_t, \mathbf{X}_t\right) = \Delta L\left(\theta_t, \mathbf{X}_t\right) + \Delta C\left(\mathbf{X}_t\right)\left(Y_t - \theta_t\right) = \boldsymbol{\alpha}'\mathbf{Z}_{t-1} + e_t + \Delta C\left(\mathbf{X}_t\right)\left(Y_t - \theta_t\right) \equiv \boldsymbol{\alpha}'\mathbf{Z}_{t-1} + \tilde{e}_t$, with $E\left[\tilde{e}_t\mathbf{Z}_{t-1}\right] = E\left[e_t\mathbf{Z}_{t-1}\right] + E\left[\Delta C\left(\mathbf{X}_t\right)\left(Y_t - \theta_t\right)\mathbf{Z}_{t-1}\right] = 0$, since $E\left[\Delta C\left(\mathbf{X}_t\right)\left(Y_t - \theta_t\right)\mathbf{Z}_{t-1}\right] = E\left[E\left[\Delta C\left(\mathbf{X}_t\right)\left(Y_t - \theta_t\right)|\mathcal{G}_{t-1}\right]\mathbf{Z}_{t-1}\right] = 0$ by part (a), and $E\left[e_t\mathbf{Z}_{t-1}\right] = 0$ under A3. Thus $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha}$ as claimed. □

Proof of Proposition 5. (a) We first obtain $E\left[\Delta L\left(Y_t, \mathbf{X}_t\right)\mathbf{Z}_{t-p}\right]$ using calculations previously presented in the proof of Proposition 3:

$$E\left[\Delta L\left(Y_t, \mathbf{X}_t\right)\mathbf{Z}_{t-p}\right] - E\left[\Delta L\left(\theta_t, \mathbf{X}_t\right)\mathbf{Z}_{t-p}\right]$$

$$= \sum_{j=1}^{J}\omega_j E\left[\Delta C\left(\mathbf{X}_t\right)\left(\tilde{\theta}_{t+j} - \theta_t\right)\mathbf{Z}_{t-p}\right]$$

$$E\left[\Delta C\left(\mathbf{X}_t\right)\left(\tilde{\theta}_{t+j} - \theta_t\right)\mathbf{Z}_{t-p}\right]$$

$$= g_0^{(j)}E\left[\Delta C\left(\mathbf{X}_t\right)\mathbf{Z}_{t-p}\right] + \sum_{i=2}^{p}g_i^{(j)}E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1-i}\mathbf{Z}_{t-p}\right]$$

$$+ \left(g_1^{(j)} - 1\right)E\left[\Delta C\left(\mathbf{X}_t\right)\theta_t\mathbf{Z}_{t-p}\right]$$

and

$$E\left[\Delta C\left(\mathbf{X}_t\right)\theta_t\mathbf{Z}_{t-p}\right] = \frac{1}{\phi_1}E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1}\mathbf{Z}_{t-p}\right]$$

$$- \frac{\phi_0}{\phi_1}E\left[\Delta C\left(\mathbf{X}_t\right)\mathbf{Z}_{t-p}\right]$$

$$- \sum_{i=2}^{p}\frac{\phi_i}{\phi_1}E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1-i}\mathbf{Z}_{t-p}\right].$$

Pulling these results together we obtain:

$$E\left[\Delta L\left(Y_t, \mathbf{X}_t\right)\mathbf{Z}_{t-p}\right] - E\left[\Delta L\left(\theta_t, \mathbf{X}_t\right)\mathbf{Z}_{t-p}\right]$$

$$= \sum_{j=1}^{J}\omega_j E\left[\Delta C\left(\mathbf{X}_t\right)\left(\tilde{\theta}_{t+j} - \theta_t\right)\mathbf{Z}_{t-p}\right]$$

$$= \sum_{j=1}^{J}\omega_j\left\{g_0^{(j)}E\left[\Delta C\left(\mathbf{X}_t\right)\mathbf{Z}_{t-p}\right]\right.$$

$$+ \sum_{i=2}^{p}g_i^{(j)}E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1-i}\mathbf{Z}_{t-p}\right] + \left(g_1^{(j)} - 1\right)$$

$$\times\left\{\frac{1}{\phi_1}E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1}\mathbf{Z}_{t-p}\right] - \frac{\phi_0}{\phi_1}E\left[\Delta C\left(\mathbf{X}_t\right)\mathbf{Z}_{t-p}\right]\right.$$

$$\left.\left. - \sum_{i=2}^{p}\frac{\phi_i}{\phi_1}E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1-i}\mathbf{Z}_{t-p}\right]\right\}\right\}$$

$$= E\left[\Delta C\left(\mathbf{X}_t\right)\mathbf{Z}_{t-p}\right]\left\{\sum_{j=1}^{J}\omega_j\left(g_0^{(j)} - g_1^{(j)}\frac{\phi_0}{\phi_1}\right) + \frac{\phi_0}{\phi_1}\right\}$$

$$+ \sum_{i=2}^{p}E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1-i}\mathbf{Z}_{t-p}\right]$$

$$\times\left\{\sum_{j=1}^{J}\omega_j\left(g_i^{(j)} - g_1^{(j)}\frac{\phi_i}{\phi_1}\right) + \frac{\phi_i}{\phi_1}\right\}$$

$$+ E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1}\mathbf{Z}_{t-p}\right]\left\{\sum_{j=1}^{J}\omega_j\frac{g_1^{(j)}}{\phi_1} - \frac{1}{\phi_1}\right\}$$

$$\equiv -\lambda_0 E\left[\Delta C\left(\mathbf{X}_t\right)\mathbf{Z}_{t-p}\right] - \sum_{i=2}^{p}\lambda_i E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1-i}\mathbf{Z}_{t-p}\right]$$

$$- \lambda_1 E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1}\mathbf{Z}_{t-p}\right].$$

Thus with $\widetilde{\Delta L\left(\theta_t, \mathbf{X}_t\right)}$ defined as in the proposition, we obtain $E\left[\widetilde{\Delta L\left(\theta_t, \mathbf{X}_t\right)}\mathbf{Z}_{t-p}\right] = E\left[\Delta L\left(\theta_t, \mathbf{X}_t\right)\mathbf{Z}_{t-p}\right].$

(b) Similar to the proof of Proposition 3(b), this part is proved by invoking a multivariate CLT for the sample mean of the loss differentials using the true volatility and all of the elements that enter into the estimated adjustment terms, $\hat{\lambda}_{i,T}$, $i = 0, 1, \ldots, p$.

This collection of elements is:

$$\mathbf{D}_t \equiv \left[\Delta L\left(\theta_t, \mathbf{X}_t\right)\mathbf{Z}'_{t-p}, \Delta C\left(\mathbf{X}_t\right)\mathbf{Z}'_{t-p}, \Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1}\mathbf{Z}'_{t-p}, \ldots,\right.$$

$$\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+J}\mathbf{Z}'_{t-p}, \ldots, \Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t-1}\mathbf{Z}'_{t-p}, \ldots,$$

$$\left.\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t-p+1}\mathbf{Z}'_{t-p}, \mathbf{Z}'_t, \tilde{\theta}_t, \tilde{\theta}_t\tilde{\theta}_{t+1}, \ldots, \tilde{\theta}_t\tilde{\theta}'_{t+2p}, \right]' \quad (37)$$

and with Assumptions A1 and A2 applied to $\mathbf{D}_t$ we have $\sqrt{T}\left(\bar{\mathbf{D}}_T - E\left[\mathbf{D}_t\right]\right) \to^d N\left(0, V_D\right)$ using Theorem 3 of Politis and Romano (1994). As in the proof of Proposition 3, the last $2p + 1$ elements of $\bar{\mathbf{D}}_T$ are sufficient to obtain estimates of P, $\mathbf{Q}_0$, and Q_1 and thus estimates of the parameters $g_i^{(j)}$, for $i = 0, 1, \ldots, p$ and $j = 1, 2, \ldots, J$. With these we obtain the estimated adjustment terms $\hat{\lambda}_{i,T}$, $i = 0, 1, \ldots, p$. Given asymptotic normality of $\bar{\mathbf{D}}_T$ and the fact that $\hat{\boldsymbol{\alpha}}_T$ is a smooth function of the elements of $\bar{\mathbf{D}}_T$, we can then apply the delta method, see Lemma 2.5 of Hayashi (2000) for example, to show asymptotic normality of $\left(\hat{\boldsymbol{\alpha}}_T - \tilde{\boldsymbol{\alpha}}\right)$, and obtain its covariance matrix. To show that $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha}$ we use the result from part (a) which provides $\tilde{\boldsymbol{\alpha}} \equiv \left(E\left[\mathbf{Z}_{t-p}\mathbf{Z}'_{t-p}\right]\right)^{-1}E\left[\mathbf{Z}_{t-p}\widetilde{\Delta L\left(\theta_t, \mathbf{X}_t\right)}\right] = \left(E\left[\mathbf{Z}_{t-p}\mathbf{Z}'_{t-p}\right]\right)^{-1}E\left[\mathbf{Z}_{t-p}\Delta L\left(\theta_t, \mathbf{X}_t\right)\right] \equiv \boldsymbol{\alpha}.$

(c) Again follows directly from Theorem 4 of Politis and Romano (1994). $\square$

References

Aït-Sahalia, Y., Mancini, L., 2008. Out of sample forecasts of quadratic variation. Journal of Econometrics 147, 17–33.

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

Andersen, T.G., 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., Christoffersen, P.F., Diebold, F.X., 2006. Volatility and correlation forecasting. In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), Handbook of Economic Forecasting. North Holland Press, Amsterdam.

Andersen, T.G., Bollerslev, T., Diebold, F.X., 2007. Roughing it up: including jump components in the measurement, modeling and forecasting of return volatility. Review of Economics and Statistics 89, 701–720.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2001a. The distribution of realized exchange rate volatility. Journal of the American Statistical Association 96, 42–55.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Ebens, H., 2001b. 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, 579–626.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2000. Great realizations. Risk 13, 105–108.

Andersen, T.G., Bollerslev, T., Huang, X., 2011a. A reduced form framework for modeling volatility of speculative prices based on realized variation measures. Journal of Econometrics 160, 176–189.

Andersen, T.G., Bollerslev, T., Meddahi, N., 2011b. Realized volatility forecasting and market microstructure noise. Journal of Econometrics 160, 220–234.

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

Bandi, F.M., Russell, J.R., 2008. Microstructure noise, realized variance, and optimal sampling. Review of Economic Studies 75, 339–369.

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

Bandi, F.M., Russell, J.R., 2006b. Comment on "realized variance and microstructure noise". Journal of Business and Economic Statistics 24, 167–173.

Bandi, F.M., Russell, J.R., 2011. Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations. Journal of Econometrics 160, 145–159.

Bandi, F.M., Russell, J.R., Yang, C., 2007. Realized volatility forecasting in the presence of time-varying noise. Working Paper. Graduate School of Business, University of Chicago.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2008. Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. Econometrica 76, 1481–1536.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2009. Realized kernels in practice. Econometrics Journal 12, 1–32.

Barndorff-Nielsen, O.E., Nielsen, B., Shephard, N., Ysusi, C., 2004. Measuring and forecasting financial variability using realised variance. In: Harvey, A., Koopman, S.J., Shephard, N. (Eds.), State Space and Unobserved Components Models: Theory and Applications. Cambridge University Press.

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, Series B 64, 253–280.

Barndorff-Nielsen, O.E., Shephard, N., 2004a. Econometric analysis of realized covariation: high frequency based covariance, regression and correlation in financial economics. Econometrica 72, 885–925.

Barndorff-Nielsen, O.E., Shephard, N., 2004b. Power and bipower variation with stochastic volatility and jumps. Journal of Financial Econometrics 2, 1–48.

Barndorff-Nielsen, O.E., Shephard, N., 2007. Variation, jumps, market frictions and high frequency data in financial econometrics. In: Blundell, R., Torsten, P., Newey, W.K. (Eds.), Advances in Economics and Econometrics. Theory and Applications, Ninth World Congress. In: Econometric Society Monographs, Cambridge University Press, pp. 328–372.

Bollerslev, T., Engle, R.F., Nelson, D.B., 1994. ARCH models. In: Engle, R.F., McFadden, D. (Eds.), Handbook of Econometrics. North Holland Press, Amsterdam.

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

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics 1–23.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–263.

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

Easley, D., O'Hara, M., 1992. Time and the process of security price adjustment. Journal of Finance 47, 577–605.

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

Engle, R.F., Patton, A.J., 2001. What good is a volatility model? Quantitative Finance 1, 237–245.

Gatheral, J., Oomen, R.C.A., 2010. Zero-intelligence realized variance estimation. Finance and Stochastics 14, 249–283.

Ghysels, E., Sinko, A., 2011. Volatility forecasting and microstructure noise. Journal of Econometrics 160, 257–271.

Giacomini, R., White, H., 2006. Tests of conditional predictive ability. Econometrica 74, 1545–1578.

Gonçalves, S., Meddahi, N., 2009. Bootstrapping realized volatility. Econometrica 77, 283–306.

Hansen, P.R., 2005. A test for superior predictive ability. Journal of Business and Economic Statistics 23, 365–380.

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

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

Hansen, P.R., Lunde, A., 2006b. Consistent ranking of volatility models. Journal of Econometrics 131, 97–121.

Hansen, P.R., Lunde, A., 2010. Estimating the persistence and the autocorrelation function of a time series that is measured with error. Working Paper.

Hansen, P.R., Lunde, A., Nason, J.M., 2010. The model confidence set. Econometrica (forthcoming).

Hayashi, F., 2000. Econometrics. Princeton University Press, New Jersey.

Huang, X., Tauchen, G., 2005. The relative contribution of jumps to total price variance. Journal of Financial Econometrics 3, 456–499.

Kalnina, I., Linton, O., 2008. Estimating quadratic variation consistently in the presence of correlated measurement error. Journal of Econometrics 147, 47–59.

Large, J., 2011. Estimating quadratic variation when quoted prices change by a constant increment. Journal of Econometrics 160, 2–11.

Maasoumi, E., McAleer, M., 2008. Realized volatility and long memory: an overview. Econometric Reviews 27, 1–9.

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

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

Meddahi, N., 2003. ARMA representation of integrated and realized variances. Econometrics Journal 6, 334–355.

Newey, W.K., West, K.D., 1987. A simple, positive semidefinite, 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.

Owens, J., Steigerwald, D., 2007. Noise reduced realized volatility: a Kalman filter approach. In: Terrell, D., Fomby, T., Hill, R.C. (Eds.), Econometric Analysis of Financial and Economic Time Series, Part A. Elsevier.

Patton, A.J., 2011. Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics 160, 246–256.

Patton, A.J., Sheppard, K., 2009a. Evaluating volatility and correlation forecasts. In: Andersen, T.G., Davis, R.A., Kreiss, J.-P., Mikosch, T. (Eds.), Handbook of Financial Time Series. Springer Verlag.

Patton, A.J., Sheppard, K., 2009b. Optimal combinations of realised volatility estimators. International Journal of Forecasting 25, 218–238.

Patton, A.J., Sheppard, K., 2010. Estimating volatility using high frequency data: an analysis across multiple markets. Work in Progress.

Patton, A.J., Timmermann, A., 2010. Generalized forecast errors, a change of measure, and forecast optimality conditions. In: Bollerslev, T., Russell, J.R., Watson, M.W. (Eds.), Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle. Oxford University Press.

Politis, D.N., Romano, J.P., 1994. The stationary bootstrap. Journal of the American Statistical Association 89, 1303–1313.

Poon, S.-H., Granger, C.W.J., 2003. Forecasting volatility in financial markets. Journal of Economic Literature 41, 478–539.

Romano, J.P., Wolf, M., 2005. Stepwise multiple testing as formalized data snooping. Econometrica 73, 1237–1282.

Tauchen, G., Zhou, H., 2011. Realized jumps on financial markets and predicting credit spreads. Journal of Econometrics 160, 102–116.

West, K.D., 1996. Asymptotic inference about predictive ability. Econometrica 64, 1067–1084.

White, H., 2000. A reality check for data snooping. Econometrica 68, 1097–1126.

White, H., 2001. Asymptotic Theory for Econometricians. Academic Press, USA.

Wright, J.H., 1999. Testing for a unit root in the volatility of asset returns. Journal of Applied Econometrics 14, 309–318.

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., 2005. A tale of two time scales: determining integrated volatility with noisy high-frequency data. Journal of the American Statistical Association 100, 1394–1411.

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