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

1. Introduction

The past decade has seen an explosion in research on volatility measurement, as distinct from volatility forecasting. 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 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.

0304-4076/$ – see front matter © 2010 Elsevier B.V. All rights reserved.

doi:10.1016/j.jeconom.2010.12.010
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, \( \theta_i \), 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:

\[
RV(m) \equiv \sum_{i=1}^{m} (p_{t_i} - p_{t_{i-1}})^2
\]

(2)

where \( p_i \) is the log-price at time \( t_i \), \( \{t_0, t_1, \ldots, t_m\} \) 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 frequencies as often as possible, see Andersen et al. (2001a) for example, as the asymptotic variance of the estimator in this case declines uniformly as \( m \rightarrow \infty \). In practice, however, the presence of autocorrelation in very high frequency prices leads the standard RV estimator to become severely biased, 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 test complements, 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 Bandi-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 RV\( IC \) estimator, which adjusts the standard RV estimator 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\( AC \) estimator, the realised kernel estimator of Bandi-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\( AC \) estimators assuming i.i.d. noise; Bandi and Russell (2006a, 2011) derive the optimal choice of \( m \) for standard RV, and the optimal ratio of \( q/m \) for RV\( AC \) 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 Bandi-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 \( m \rightarrow \infty \), 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 \(\theta\). I assume that \(\theta\) 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 \(\theta\) is a scalar; I discuss the extension to vector (or matrix) target variables in the conclusion. The estimators of \(\theta\) are denoted \(\hat{\theta}_i\), \(i = 1, 2, \ldots, k\). 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, \(X_{\hat{\theta}_i}\), to the target variable, \(\theta\). Two popular (pseudo-)distance measures in the volatility literature are MSE and QLIKE:

\[
\text{MSE} \quad L(\theta, X) = (\theta - X)^2
\]

\[
\text{QLIKE} \quad L(\theta, X) = \frac{\theta}{X} - \log \left(\frac{\theta}{X}\right) - 1.
\]

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:

\[
L(\theta, X) = \tilde{C}(X) - \tilde{C}(\theta) + C(\theta)(\theta - X)
\]

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:

\[
\Delta L(\cdot, X_i) \equiv [L(\cdot, X_{i1}) - L(\cdot, X_{i2}), \ldots, L(\cdot, X_{it}) - L(\cdot, X_{it})]^\prime
\]

where \(X_i = [X_{i1}, \ldots, X_{ik}]^\prime\).

Throughout, variables denoted with a “\(\sim\)“ below are the bootstrap samples of the original variables obtained from the stationary bootstrap, \(\bar{P}\) 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 \(\bar{\theta}_i\) to denote conditional variance in part (a) and quadratic variation in part (b).

**Proposition 1.** Let \(\theta_t\) 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 up to time \(t\), and let \(\mathcal{F}_b \subset \mathcal{F}_t\) be the information set available to the econometrician at time \(t\). Let \((X_{1t}, X_{2t})\) be two estimators of \(\theta_t\), and let \(\hat{\theta}_i\) be the proxy for \(\theta_t\).

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

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

\[
\Rightarrow E \left[ L \left( \hat{\theta}_i, X_{1t} \right) \right] \leq E \left[ L \left( \hat{\theta}_i, X_{2t} \right) \right].
\]

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

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

\[
\Rightarrow E \left[ L \left( \hat{\theta}_i, X_{1t} \right) \right] \leq E \left[ L \left( \hat{\theta}_i, X_{2t} \right) \right].
\]

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 \(\hat{\theta}_i\) 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, \(E [\Delta L (\theta_t, X_i)]\), by exploiting some well-known empirical properties of the behaviour of \(\theta_t\) and by making use of a (function of a) proxy for \(\theta_t\), denoted \(\tilde{\theta}_i\). 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.** \(\tilde{\theta}_i = \theta_t + \nu_i\), with \(E [\nu_i | \mathcal{F}_{t-1}, \theta_t] = 0, \text{a.s.}\)

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 Barndoff-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.** \(Y_t = \sum_{i=1}^{\infty} \omega_i \tilde{\theta}_{t+i}, \text{where } 1 \leq j < \infty, \omega_0 \geq 0 \forall i\) and \(\sum_{i=0}^{\infty} \omega_i = 1.\)
In the propositions below I consider using a convex combination of leads of $\theta_t$, as in Assumption P2, the simplest special case of which is just a one-period lead (and so $Y_t = \theta_{t+1}$). Using leads of the proxy is important for breaking the correlated measurement error problem, which makes it possible to overcome the problems identified in Proposition 1. $Y_t$ is thus interpretable as an instrument for $\theta_t$. 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 $\text{Corr} \left[ Y_t, \theta_t \right] \neq 0$; some more structure is required. I obtain results in this application by considering two alternative approximations of the conditional mean of $\theta_t$.

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 the conditional mean of the proxy is important for breaking the correlated measurement error problem, which makes it possible to overcome the problems identified in Proposition 1. $Y_t$ is thus interpretable as an instrument for $\theta_t$. 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 $\text{Corr} \left[ Y_t, \theta_t \right] \neq 0$; some more structure is required. I obtain results in this application by considering two alternative approximations of the conditional mean of $\theta_t$.

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 $E [ L (\theta_t, X_t) ] = E [ L (Y_t, X_t) ]$ for any vector of RV estimators, $X_t$, and any $L$ such that these expectations exist.

(b) If we further assume A1 and A2 in the Appendix, then:

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

where $\Omega_1$ is given in the proof.

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

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

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, $\theta_t$, 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. 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, $\theta_t$, to follow a stationary AR(p) process. The work of Meddahi (2003) and Barndoff-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. $\theta_t = \phi_0 + \sum_{i=1}^{\infty} \phi_i \theta_{t-i} + \eta_t$, with $E [ \eta_t | \mathcal{F}_{t-1} ] = 0$ a.s., $\theta_t > 0$ a.s., $\phi_1 \neq 0$, the matrix $\Psi$ defined in Eq. (36) is invertible, and $\Phi = \begin{bmatrix} \phi_1 & \cdots & \phi_p \end{bmatrix}$ is such that $\theta_t$ 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
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 $v_{t-j}$ 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 $Q_0 = \left[\phi_0, \Phi_0\right], Q_1 = \left[\phi_0, \Phi_0, 0\right], P = \left[0, \Phi_p\right]$, where $\Phi_p$ is a $p \times 1$ vector of zeros. Then:

(a) $E\left[\Delta L (\theta_t, X_t)\right] = E\left[\Delta L (Y_t, X_t)\right] - \beta$

where

$$\beta = E\left[\Delta C (X_t)\right] \sum_{j=1}^{J} \omega_j g^{(i)}_j / S^{(i)}_1$$

$$\alpha \sum_{j=1}^{J} \omega_j \left(1 - 1/\alpha g^{(i)}_j\right) E\left[\Delta C (X_t) \tilde{\theta}_{t+j}\right]$$

$$\sum_{j=1}^{J} \omega_j \left(1 - 1/\alpha g^{(i)}_j\right)^2 E\left[\Delta C (X_t) \tilde{\theta}_{t+j}\right]$$

for any vector of RV estimators, $X_t$, and any L such that these expectations exist. The variable $g^{(i)}_j$ is defined as the first element of the vector $(1 - (p-1)Q_1)^j (1 - p^{-1}Q_1)^{1-p} Q_0$, and $g^{(i)}_j$ is defined as (1, i) element of the matrix $(P^{-1}Q_j)^j$.

(b) If we further assume A1 and A2 in the Appendix hold for the series $B_t$, defined in Eq. (35), then:

$$\sqrt{T} \left( \frac{1}{T} \sum_{t=1}^{T} \Delta L (Y_t, X_t) - \hat{\beta}_t \right) \rightarrow_d N (0, \Omega_T) \Rightarrow \text{as } T \rightarrow \infty$$

where

$$\hat{\beta}_t = \left( \frac{1}{T} \sum_{t=1}^{T} \Delta C (X_t) \right) \left( \sum_{j=1}^{J} \omega_j g^{(i)}_j / S^{(i)}_1 \right)$$

$$\alpha \sum_{j=1}^{J} \omega_j \left(1 - 1/\alpha g^{(i)}_j\right) 1 - 1/j \sum_{i=1}^{J} \Delta C (X_t) \tilde{\theta}_{t+j}$$

$$\sum_{j=1}^{J} \omega_j \left(1 - 1/\alpha g^{(i)}_j\right)^2 1 - 1/j \sum_{i=1}^{J} \Delta C (X_t) \tilde{\theta}_{t+j}$$

where $g^{(i)}_j, i = 0, 1, \ldots, p; j = 1, 2, \ldots, J$ 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:

$$\sup_{x} \left| \frac{1}{T} \sum_{t=1}^{T} \Delta L (Y_t, X_t) - \hat{\beta}_t \right| \leq z$$

$$\left| \frac{-1}{T} \sum_{t=1}^{T} \Delta L (Y_t, X_t) + \hat{\beta}_t \right| \leq z$$

$$\left| - P \left( \frac{1}{T} \sum_{t=1}^{T} \Delta L (Y_t, X_t) - \hat{\beta}_t \right) \leq z \right| \rightarrow 0, \text{ as } T \rightarrow \infty.$$
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, \( \theta_i \), to follow a stationary AR\(p\) process. The proposition below shows that the AR approximation can be accomplished by using an adjusted dependent variable in the GW-type regression. That is, the infeasible regression is again:

\[
\Delta L(\theta_i, X_i) = \alpha' Z_{t-p} + e_t \tag{11}
\]

while the adjusted regression becomes:

\[
\Delta L(\bar{\theta}_i, X_i) = \bar{\alpha}' Z_{t-p} + \bar{e}_t. \tag{12}
\]

Note that the variable \( Z_t \) 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 \( Z_{t-p} \). Under the random walk approximation the adjusted dependent variable is simply \( \Delta L(\bar{\theta}_i, X_i) = \Delta L(Y_t, X_i) \), 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 \( Y_t = \hat{\theta}_{t+1} \)) we have:

\[
\Delta L(\bar{\theta}_i, X_i) = \Delta L\left(\hat{\theta}_{t+1}, X_i\right) - \frac{\phi_0}{\phi_1} \Delta C(X_i) + \frac{1 - \phi_1}{\phi_1} \Delta C(X_i) \hat{\theta}_{t+1}. \tag{13}
\]

This adjusted dependent variable is constructed such that \( \bar{\alpha} = \alpha \), and thus estimating Eq. (12) by OLS yields a consistent estimator of the unknown true parameter \( \alpha \). (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:

\[
\Delta L(\bar{\theta}_i, X_i) = \tilde{\alpha}' \tilde{Z}_{t-p} + \tilde{e}_t \tag{14}
\]

where

\[
\Delta L(\bar{\theta}_i, X_i) = \Delta L\left(\hat{\theta}_{t+1}, X_i\right) - \frac{\phi_0}{\phi_1} \Delta C(X_i) \hat{\theta}_{t+1} + \frac{1 - \phi_1}{\phi_1} \Delta C(X_i) \hat{\theta}_{t+1} \tag{15}
\]

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 \( \hat{\theta}_i \), 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’.** \( X_i \) is conditionally independent of \( \nu_{t-j} \) given \( F_{t-j} \), 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 \( \theta_0, \theta_1, \theta_p \) and \( g_i^{(j)} \) be defined as in Proposition 3. Finally, assume that \( \Delta L(\theta_i, X_i) \) is a scalar, and define:

\[
\Delta L(\bar{\theta}_i, X_i) = \Delta L(Y_t, X_i) + \lambda_0 \Delta C(X_i) + \lambda_1 \Delta C(X_i) \hat{\theta}_{t+1} + \sum_{i=2}^{p} \lambda_i \Delta C(X_i) \hat{\theta}_{t+1-i}
\]

where

\[
\lambda_1 = \frac{1}{\phi_1} - \frac{\phi_0}{\phi_1} \frac{\sum_{j=1}^{p} \phi_j^{(j)} \phi_1}{\phi_1}, \quad \text{and} \quad \lambda_i = -\frac{\phi_i}{\phi_1} \left( \phi_0^{(j)} - \phi_j^{(j)} \phi_1 \right), \quad i = 0, 2, 3, \ldots, p
\]

and

\[
\Delta L(\bar{\theta}_i, X_i) = \Delta L(Y_t, X_i) + \lambda_0 \Delta C(X_i) + \sum_{i=2}^{p} \lambda_i \Delta C(X_i) \hat{\theta}_{t+1-i}
\]

where \( \lambda_i \), \( i = 0, 1, \ldots, p \) are the values of \( \lambda_i \) based on estimated values for \( \phi_0 \) and \( \phi_0^{(j)} \). Then:

(a) \( E \left[ \Delta L(\bar{\theta}_i, X_i) Z_{t-p} \right] = E \left[ \Delta L(\theta_i, X_i) Z_{t-p} \right] \)

for any \( Z_{t-p} \in F_{t-p} \).

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

\[ \sqrt{T} \left( \hat{\alpha}_T - \alpha \right) \to^d N \left( 0, \Omega_2 \right) \text{ as } T \to \infty. \]

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

\[ \sup_z \left[ P \left[ \| \hat{\alpha}_T - \alpha \| \leq z \right] - P \left[ \| \alpha - \alpha \| \leq z \right] \right] \to 0. \text{ as } T \to \infty. \]

As in Proposition 3, the AR assumption introduces additional terms to be estimated in order to consistently estimate \( E \left[ \Delta L(\theta_i, X_i) Z_{t-p} \right] \). 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 \( Z_{t-p} \) 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:

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

\[
d \nu^2(t) = 0.035 \left( 0.636 - \nu^2(t) \right) d(t) + 0.144 \nu^2(t) dW_1(t). \tag{17}
\]

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:

\[
d \log \nu^2(t) = -0.0136 \left( 0.8382 + \log \nu^2(t) \right) d(t) + 0.1148 dW_1(t). \tag{18}
\]

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

\[
d \log P(t) = 0.030 d(t) + \nu(t) \left( -0.30 dW_1(t) - 0.30 dW_2(t) + \sqrt{1 - 2 \times (0.30)^2} dW_2(t) \right) \tag{16}
\]

\[
d \nu^2(t) = \exp \left[ -1.2 + 0.04 \nu^2(t) + 1.5 \nu^2(t) \right] \tag{18}
\]

\[
d \nu^2(t) = -0.00137 \nu^2(t) dW_1(t) + 1 + 0.25 \nu^2(t) dW_2(t). \tag{18}
\]

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:

$$
\log P(t) = \log P^0(t) + \xi(t)
$$

(19)

Following Aït-Sahalia et al. (2005) and Huang and Tauchen (2005), I set $\sigma_\xi^2$ 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%:

$$
\frac{2\sigma_\xi^2}{V[r_{30min}^5] + 2\sigma_\xi^2} = 0.20
$$

(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 method is investigated via the following experiment. For specified cases where both the random walk and AR($p$) processes 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.

As in the simulation for tests of unconditional accuracy, I choose $\omega$ and $\sigma_\omega^2$ such that $\sigma_\omega^2/V[QV_t] = \sigma_\omega^2/V[QV_t] = 0.1$ and $\omega = \rho \sigma_\omega^2$. The corresponding results for the MSE distance measure are similar and available upon request.

As in the simulation for tests of unconditional accuracy, I choose $\omega$ and $\sigma_\omega^2$ such that $\sigma_\omega^2/V[QV_t] = \sigma_\omega^2/V[QV_t] = 0.1$ and $\omega = \rho \sigma_\omega^2$. The corresponding results for the MSE distance measure are similar and available upon request.

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.

As in the simulation for tests of unconditional accuracy, I choose $\omega$ and $\sigma_\omega^2$ such that $\sigma_\omega^2/V[QV_t] = \sigma_\omega^2/V[QV_t] = 0.1$ and $\omega = \rho \sigma_\omega^2$. The corresponding results for the MSE distance measure are similar and available upon request.

As in the simulation for tests of unconditional accuracy, I choose $\omega$ and $\sigma_\omega^2$ such that $\sigma_\omega^2/V[QV_t] = \sigma_\omega^2/V[QV_t] = 0.1$ and $\omega = \rho \sigma_\omega^2$. The corresponding results for the MSE distance measure are similar and available upon request.
where \( \tilde{\theta}_1 \) 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>
<thead>
<tr>
<th>Table 1</th>
<th>Finite-sample size and power of unconditional accuracy tests.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T )</td>
<td>QV ( ^* )</td>
</tr>
<tr>
<td>500</td>
<td>500 2500</td>
</tr>
<tr>
<td>0.10</td>
<td>0.02 0.02</td>
</tr>
<tr>
<td>0.15</td>
<td>0.30 0.53</td>
</tr>
<tr>
<td>0.20</td>
<td>0.50 0.79</td>
</tr>
<tr>
<td>0.50</td>
<td>0.75 0.99</td>
</tr>
<tr>
<td>1.00</td>
<td>0.84 0.99</td>
</tr>
</tbody>
</table>

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 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 (\( \gamma = \sigma_2^2/V[QV] \)) than the first (\( \sigma_1^2/V[QV] = 0.10 \)).

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Finite-sample size and power of conditional accuracy tests, slope coefficient t-test.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T )</td>
<td>QV ( ^* )</td>
</tr>
<tr>
<td>500</td>
<td>500 2500</td>
</tr>
<tr>
<td>0.00</td>
<td>0.02 0.03</td>
</tr>
<tr>
<td>0.10</td>
<td>0.06 0.15</td>
</tr>
<tr>
<td>0.20</td>
<td>0.13 0.38</td>
</tr>
<tr>
<td>0.40</td>
<td>0.23 0.74</td>
</tr>
<tr>
<td>0.80</td>
<td>0.17 0.53</td>
</tr>
</tbody>
</table>

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 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 (\( \gamma = \sigma_2^2/V[QV] \)) than the first (\( \sigma_1^2/V[QV] = 0.10 \)).
Table 3

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

<table>
<thead>
<tr>
<th>( \lambda )</th>
<th>RW AR ( \lambda )</th>
<th>RW AR ( \lambda )</th>
<th>RW AR ( \lambda )</th>
<th>RW AR ( \lambda )</th>
<th>RW AR ( \lambda )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{500}{500} ) ( \frac{2500}{2500} ) ( \frac{500}{500} )</td>
<td>( \frac{2500}{2500} ) ( \frac{500}{500} )</td>
<td>( \frac{2500}{2500} ) ( \frac{500}{500} )</td>
<td>( \frac{2500}{2500} ) ( \frac{500}{500} )</td>
<td>( \frac{2500}{2500} ) ( \frac{500}{500} )</td>
<td>( \frac{2500}{2500} ) ( \frac{500}{500} )</td>
</tr>
<tr>
<td>0.00</td>
<td>0.02</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>0.10</td>
<td>0.46</td>
<td>0.62</td>
<td>0.44</td>
<td>0.61</td>
<td>0.25</td>
</tr>
<tr>
<td>0.20</td>
<td>0.75</td>
<td>0.92</td>
<td>0.75</td>
<td>0.93</td>
<td>0.47</td>
</tr>
<tr>
<td>0.40</td>
<td>0.96</td>
<td>1.00</td>
<td>0.95</td>
<td>1.00</td>
<td>0.56</td>
</tr>
<tr>
<td>0.80</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.78</td>
</tr>
</tbody>
</table>

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 \(-\lambda \times QV_{t-1}\).

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. This sample period covers several distinct periods:

1. 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. This sample period covers several distinct periods:

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 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, using the QLIKE distance measure presented in Eq. (4). The first figure uses the random walk (RW) approximation for the dynamics in QV, the second uses a first-order AR approximation. 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.

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.

The conclusion from these pictures is that there are clear gains in 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:

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

versus

$$H_{1}^{(s)} : E \left[ L(\theta_{t}, X_{t,0}) \right] > E \left[ L(\theta_{t}, X_{t,s}) \right]$$

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 significantly worse.
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.

<table>
<thead>
<tr>
<th>Sampling frequency</th>
<th>RW approximation</th>
<th>AR approximation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trades</td>
<td>Quotes</td>
</tr>
<tr>
<td></td>
<td>Calendar</td>
<td>Tick</td>
</tr>
<tr>
<td>1 s</td>
<td>✓✓✓✓</td>
<td>–</td>
</tr>
<tr>
<td>2 s</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓</td>
</tr>
<tr>
<td>5 s</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓</td>
</tr>
<tr>
<td>15 s</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓</td>
</tr>
<tr>
<td>30 s</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓</td>
</tr>
<tr>
<td>1 min</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓</td>
</tr>
<tr>
<td>2 min</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓</td>
</tr>
<tr>
<td>5 min</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓</td>
</tr>
<tr>
<td>15 min</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓</td>
</tr>
<tr>
<td>30 min</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓</td>
</tr>
<tr>
<td>1 h</td>
<td>✓✓✓✓</td>
<td>✓✓✓✓</td>
</tr>
<tr>
<td>1 day</td>
<td>⋆</td>
<td>–</td>
</tr>
</tbody>
</table>

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.

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.

<table>
<thead>
<tr>
<th>Sampling frequency</th>
<th>RW approximation</th>
<th>AR approximation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trades</td>
<td>Quotes</td>
</tr>
<tr>
<td></td>
<td>Calendar</td>
<td>Tick</td>
</tr>
<tr>
<td>1 s</td>
<td>~~~~</td>
<td>~</td>
</tr>
<tr>
<td>2 s</td>
<td>~</td>
<td>X X X ~</td>
</tr>
<tr>
<td>5 s</td>
<td>~</td>
<td>X X X ~</td>
</tr>
<tr>
<td>15 s</td>
<td>~</td>
<td>X X X ~</td>
</tr>
<tr>
<td>30 s</td>
<td>~</td>
<td>X X X ~</td>
</tr>
<tr>
<td>1 min</td>
<td>~</td>
<td>X X X ~</td>
</tr>
<tr>
<td>2 min</td>
<td>~</td>
<td>X X X ~</td>
</tr>
<tr>
<td>5 min</td>
<td>~</td>
<td>X X X ~</td>
</tr>
<tr>
<td>15 min</td>
<td>~</td>
<td>X X X ~</td>
</tr>
<tr>
<td>30 min</td>
<td>~</td>
<td>X X X ~</td>
</tr>
<tr>
<td>1 h</td>
<td>~</td>
<td>X X X ~</td>
</tr>
<tr>
<td>2 h</td>
<td>~</td>
<td>X X X ~</td>
</tr>
<tr>
<td>1 day</td>
<td>~</td>
<td>X X X ~</td>
</tr>
</tbody>
</table>

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 different are marked with a ‘‘’). 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. 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” 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.

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 RVs 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 RVs in the full sample. It is noteworthy that in the latter sub-sample (2004–2007) all four of the more sophisticated estimators significantly beat RVs, 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 over-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 \( g_{t-1} \), that is:

\[
H^*_G : E [ L (\hat{\theta}_t, X_{t0}) | g_{t-1} ] - E [ L (\tilde{\theta}_t, X_{t0}) | g_{t-1} ] = 0
\]

a.s. \( t = 1, 2, \ldots \)

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

\[
L (\hat{\theta}_t, X_{t0}) - L (\tilde{\theta}_t, X_{t0}) = \beta_0 + \beta_1Z_{t-1} + \epsilon_t
\]

where \( Z_{t-1} \in g_{t-1} \), and then test the necessary conditions:

\[
H_0 : \beta_0 = \beta_1 = 0
\]

versus

\[
H_1 : \beta_i \neq 0 \quad \text{for some } i = 0, 1.
\]

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.
Hansen and present results in Diebold 20 Table 7
Proposition 5

Residuals, up to the tenth lag, yield a parametric “pure jump” model of high frequency asset prices, assumptions about the trading arrival process. For example, in the present paper allow us to avoid making any specific assumptions on the arrival rate of trades, while the methods presented in this paper allow us to avoid making any specific assumptions about the arrival rate of trades, the placing and removing of limit and market orders.

Calendar-time sampling and tick-time sampling are equivalent for the 1-s and 1-day sampling frequencies. In general, if the trade arrival rate is irregular, 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

below the parameter estimates18:

\[
L (Y_t, RV_t^{\text{daily}}) - L (Y_t, RV_t^{\text{5 min}}) = 33.67 + \epsilon_t \quad (31)
\]

\[
L (Y_t, RV_t^{\text{daily}}) - L (Y_t, RV_t^{\text{daily}}) = 24.94 + 17.85 Z_{t-1} + \epsilon_t \quad (32)
\]

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, \(RV_t^{\text{daily}}\), are less accurate on average than \(RV_t^{\text{5 min}}\) 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 obtained19:

\[
L (Y_t, RV_t^{\text{daily}}) - L (Y_t, RV_t^{\text{5 min}}) = 33.54 + \epsilon_t \quad (33)
\]

\[
L (Y_t, RV_t^{\text{daily}}) - L (Y_t, RV_t^{\text{daily}}) = 19.93 + 27.76 Z_{t-1} + \epsilon_t \quad (34)
\]

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 averaging 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 AR 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 VV) 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

Notes: This table presents the results of comparisons of the accuracy of seven estimators of VV: 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 Hưng), and realised range-based variance (RRV). Two approximations for the dynamics of VV 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 results in average QVKE accuracy, relative to \(RV_t^{\text{5 min}}\) are reported in the first column of each panel, with negative (positive) \(Z\)-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 0.011.

Table 6
Comparing \(RV_t^{\text{5 min}}\) with more sophisticated estimators.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>RW approximation</th>
<th>AR approximation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\text{Avg } \Delta L)</td>
<td>(\text{t-statistics on } \Delta L)</td>
</tr>
<tr>
<td></td>
<td>99–07</td>
<td>96–99</td>
</tr>
<tr>
<td>(RV_t^{1 s})</td>
<td>-0.01</td>
<td>-1.02</td>
</tr>
<tr>
<td>(RV_t^{5 \text{ min}})</td>
<td>0.00</td>
<td>⋆</td>
</tr>
<tr>
<td>(RV_t^{\text{1 day}})</td>
<td>29.66</td>
<td>9.77</td>
</tr>
<tr>
<td>TSRV</td>
<td>-0.00</td>
<td>-0.09</td>
</tr>
<tr>
<td>MSRV</td>
<td>-0.00</td>
<td>-0.29</td>
</tr>
<tr>
<td>RK Hưng</td>
<td>-0.01</td>
<td>-2.14</td>
</tr>
<tr>
<td>RRV</td>
<td>-0.02</td>
<td>-3.41</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Sampling frequency</th>
<th>RV approximation</th>
<th>AR approximation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unconditional</td>
<td>Conditional</td>
</tr>
<tr>
<td></td>
<td>(t-stat)</td>
<td>(t-stat)</td>
</tr>
<tr>
<td>2 s</td>
<td>0.01* (13.92)</td>
<td>0.08* (13.32)</td>
</tr>
<tr>
<td></td>
<td>(–2.73)</td>
<td>(–2.67)</td>
</tr>
<tr>
<td>5 s</td>
<td>0.02* (12.55)</td>
<td>0.07* (12.32)</td>
</tr>
<tr>
<td></td>
<td>(–2.20)</td>
<td>(–2.40)</td>
</tr>
<tr>
<td>15 s</td>
<td>0.01* (4.83)</td>
<td>–0.15 (4.93)</td>
</tr>
<tr>
<td></td>
<td>(–0.80)</td>
<td>(–0.83)</td>
</tr>
<tr>
<td>30 s</td>
<td>–0.00 (–1.15)</td>
<td>–0.02 (–0.40)</td>
</tr>
<tr>
<td></td>
<td>(–2.80)</td>
<td>(–2.30)</td>
</tr>
<tr>
<td>1 min</td>
<td>–0.00* (–0.00)</td>
<td>0.07* (–0.00)</td>
</tr>
<tr>
<td></td>
<td>(–2.80)</td>
<td>(–2.90)</td>
</tr>
<tr>
<td>2 min</td>
<td>–0.01* (–3.81)</td>
<td>0.02 (–0.57)</td>
</tr>
<tr>
<td></td>
<td>(–2.80)</td>
<td>(–2.30)</td>
</tr>
<tr>
<td>5 min</td>
<td>–0.02* (–6.20)</td>
<td>–0.01 (–3.10)</td>
</tr>
<tr>
<td></td>
<td>(–2.20)</td>
<td>(–2.20)</td>
</tr>
<tr>
<td>15 min</td>
<td>–0.06* (–6.50)</td>
<td>0.17 (1.01)</td>
</tr>
<tr>
<td></td>
<td>(–2.80)</td>
<td>(–2.30)</td>
</tr>
<tr>
<td>30 min</td>
<td>–0.06* (–4.92)</td>
<td>0.70* (1.63)</td>
</tr>
<tr>
<td></td>
<td>(–2.20)</td>
<td>(–2.20)</td>
</tr>
<tr>
<td>1 h</td>
<td>–0.25* (–3.59)</td>
<td>1.97 (1.63)</td>
</tr>
<tr>
<td></td>
<td>(–2.75)</td>
<td>(–2.20)</td>
</tr>
<tr>
<td>2 h</td>
<td>–1.00* (–3.04)</td>
<td>10.66 (7.94)</td>
</tr>
<tr>
<td></td>
<td>(–2.75)</td>
<td>(–2.75)</td>
</tr>
</tbody>
</table>

Notes: This table presents the estimated difference in average distance of tick-time and calendar-time RV estimators, $L(\hat{Y}_t, RV^{(tick)}) - L(\hat{Y}_t, RV^{(cal)})$, 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.

<table>
<thead>
<tr>
<th>Sampling frequency</th>
<th>RV approximation</th>
<th>AR approximation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unconditional</td>
<td>Conditional</td>
</tr>
<tr>
<td></td>
<td>(t-stat)</td>
<td>(t-stat)</td>
</tr>
<tr>
<td>1 s</td>
<td>0.11* (0.20)</td>
<td>0.33* (1.73)</td>
</tr>
<tr>
<td></td>
<td>(–9.31)</td>
<td>(–9.84)</td>
</tr>
<tr>
<td>2 s</td>
<td>0.11* (0.20)</td>
<td>0.34* (1.20)</td>
</tr>
<tr>
<td></td>
<td>(–9.31)</td>
<td>(–9.84)</td>
</tr>
<tr>
<td>5 s</td>
<td>0.11* (10.00)</td>
<td>0.34* (12.30)</td>
</tr>
<tr>
<td></td>
<td>(–9.31)</td>
<td>(–9.84)</td>
</tr>
<tr>
<td>15 s</td>
<td>0.11* (11.89)</td>
<td>0.30* (11.73)</td>
</tr>
<tr>
<td></td>
<td>(–9.31)</td>
<td>(–9.84)</td>
</tr>
<tr>
<td>30 s</td>
<td>0.08* (12.56)</td>
<td>0.24* (13.70)</td>
</tr>
<tr>
<td></td>
<td>(–10.20)</td>
<td>(–10.20)</td>
</tr>
<tr>
<td>1 min</td>
<td>0.06* (1.10)</td>
<td>0.15* (1.09)</td>
</tr>
<tr>
<td></td>
<td>(–10.30)</td>
<td>(–10.30)</td>
</tr>
<tr>
<td>2 min</td>
<td>0.04* (1.10)</td>
<td>0.15* (1.09)</td>
</tr>
<tr>
<td></td>
<td>(–10.30)</td>
<td>(–10.30)</td>
</tr>
<tr>
<td>5 min</td>
<td>0.03* (0.72)</td>
<td>0.10* (0.60)</td>
</tr>
<tr>
<td></td>
<td>(–10.90)</td>
<td>(–10.90)</td>
</tr>
<tr>
<td>15 min</td>
<td>0.03* (0.33)</td>
<td>0.07* (0.33)</td>
</tr>
<tr>
<td></td>
<td>(–4.18)</td>
<td>(–4.18)</td>
</tr>
<tr>
<td>30 min</td>
<td>0.03* (0.33)</td>
<td>0.07* (0.33)</td>
</tr>
<tr>
<td></td>
<td>(–4.18)</td>
<td>(–4.18)</td>
</tr>
<tr>
<td>1 h</td>
<td>–0.04 (–0.77)</td>
<td>–0.07 (–0.90)</td>
</tr>
<tr>
<td></td>
<td>(–1.47)</td>
<td>(–1.47)</td>
</tr>
<tr>
<td>2 h</td>
<td>–0.19 (–1.47)</td>
<td>1.96 (1.47)</td>
</tr>
<tr>
<td></td>
<td>(–1.47)</td>
<td>(–1.47)</td>
</tr>
<tr>
<td>1 day</td>
<td>1.98 (0.41)</td>
<td>10.70 (0.41)</td>
</tr>
<tr>
<td></td>
<td>(0.41)</td>
<td>(0.41)</td>
</tr>
</tbody>
</table>

Notes: This table presents the estimated difference in average distance of quote-price and trade-price RV estimators, $L(\hat{Y}_t, RV^{(qt)}) - L(\hat{Y}_t, RV^{(tr)})$, 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 trade observations 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.
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.

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 (2002) 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

I thank the Editor, Ron Gallant, three anonymous referees, and Alan Bester, Tim Bollerslev, Peter Hansen, Nour Meddahi, Roel Oomen, Neil Shephard, Kevin Sheppard and seminar participants at Athens University of Economics & Business, Cambridge, Chicago, CREATES in Aarhus, Duke, Essex, Federal Reserve Bank of St. Louis, Imperial College London, Keele, Lugano, Queensland University of Technology, Tinbergen Institute Amsterdam, University College Dublin, the London-Oxford Financial Econometrics workshop, the STIE workshop in Stanford, and the workshop on Model Evaluation and Predictive Ability in Paris for helpful comments and suggestions. Runquan Chen and Ben Carlston provided excellent research assistance. Financial support from the Leverhulme Trust under grant F/0004/AF is gratefully acknowledged.

Appendix. Proofs

Additional assumptions used in parts of the proofs below:

Let \( A_i = \left[ \Delta \left( \theta, X_i \right) \right] \), let \( A_{j,k} \) denote the sample mean of \( A_i \), and let \( A_{j,k} \) denote the \( j \)th element of \( A_i \).

Assumption A1. \( E \left[ A_{j,k} \right] \) is positive definite.

Assumption A2. \( \{ A_i \} \) is a-mixing of size \(-3 (6 + \varepsilon) / \varepsilon \).

Assumption A3. \( E \left[ Z_{t-1} \right] \) is positive definite.

Assumption A4a. \( \left( Z_{t-1}, \tilde{e}_t \right) \) is a-mixing of size \((2 + \varepsilon) / \varepsilon \) for some \( \varepsilon > 0 \).

Assumption A4b. \( E \left[ Z_{t-1}, \tilde{e}_t^2 \right] < \infty \) for \( i = 1, 2, \ldots, q \) and all \( t \).

Assumption A4c. \( V_t = \sum_{t=1}^T Z_{t-1} \tilde{e}_t \) is uniformly positive definite.

Assumption A4d. \( E \left[ Z_{t-1} \tilde{e}_t^2 \right] < \infty \) for some \( \delta > 0 \) and all \( i = 1, 2, \ldots, q \) and all \( t \).

Assumption A4e. \( M_T = \sum_{t=1}^T Z_{t-1} \tilde{e}_t \) 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 \( L (\tilde{\theta}, X_{it}) \) around \( (\tilde{\theta}, X_{it}) \):

\[
L (\tilde{\theta}, X_{it}) = L (\tilde{\theta}, X_{it}) + \frac{\partial L (\tilde{\theta}, X_{it})}{\partial \theta} (\tilde{\theta} - \tilde{\theta}_i) + \frac{1}{2} \frac{\partial^2 L (\tilde{\theta}, X_{it})}{\partial \theta^2} (\tilde{\theta} - \tilde{\theta}_i)^2
\]

\[
= L (\tilde{\theta}, X_{it}) + \left( C (X_{it}) - C (\tilde{\theta}_i) \right) (\tilde{\theta} - \tilde{\theta}_i) - \frac{1}{2} C' (\tilde{\theta}_i) (\tilde{\theta} - \tilde{\theta}_i)^2
\]

where \( \tilde{\theta}_i = \lambda_i \tilde{\theta}_i + (1 - \lambda_i) \tilde{\theta}_i \) for some \( \lambda_i \in [0, 1] \), 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 \( X_{it} \). In volatility forecasting applications, \( \tilde{\theta}_i \) is the conditional variance and so \( \tilde{\theta}_i \in \mathcal{F}_{t-1} \), and \( X_{it} \) is a volatility forecast, and so \( X_{it} \in \mathcal{F}_{t-1} \). In that case, this allows

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

by the unbiasedness of \( \tilde{\theta}_i \) for \( \tilde{\theta}_i \) conditional on \( \mathcal{F}_{t-1} \). Using the law of iterated expectations we obtain

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

and thus

\[
E \left[ \Delta L (\tilde{\theta}, X_{it}) \right] = E \left[ \Delta L (\tilde{\theta}, X_{it}) \right].
\]

(b) When \( X_{it} \) is a realised volatility estimator and \( \tilde{\theta}_i \) is the integrated variance or quadratic variation we have \( \tilde{\theta}_i \in \mathcal{F}_{t-1} \) and
\((X_t, \tilde{\theta}_t) \in \mathcal{F}_t\), which means we cannot employ the above reasoning directly. If we could assume that \(\text{Corr}[C(X_t) - C(\theta_t) \cdot \tilde{\theta}_t - \theta_t | \mathcal{F}_{t-1}] = 0 \, \forall i\), in addition to \(E[\tilde{\theta}_t | \mathcal{F}_t] = \theta_t\), then we would have

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

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

**Proof of Proposition 2.** (a) See the proof of part (a) Proposition 4 and set \(g^{\cdot \cdot}_t - \theta_t \) to be the trivial information set.

(b) Note that

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

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

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

\[
\sqrt{T} \left( \mathbf{A}_t - E [\mathbf{A}_t] \right) \rightarrow^d N (0, \Omega_T)
\]

where \(\Omega_T \equiv \mathbf{V}_T \mathbf{A}_t \mathbf{V}_T\). It should be noted that Assumptions A1 and A2 can hold despite the random walk Assumption T1, if \(\theta_t\) and \(X_t\) 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

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

where \(\beta = E [\Delta C(\mathbf{X}_t) Y_t - \theta_t] \equiv \sum_{j=1}^{p} \omega_j E [\Delta C(\mathbf{X}_t) (E_t [\theta_{t+j}] - \theta_t)]\). Under P1 and P2.

Allowing for \(j > 1\) requires computing \((j > 1)\)-step ahead forecasts from an AR(p) process, \(E_t [\theta_{t+j}]\). This is simplified by using the companion form for the AR(p) process governing \(\theta_t\):

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

 redefine as

\[
PZ_t = Q_0 + Q_1 Z_{t-1} + V_t
\]

so

\[
Z_t = P^{-1} Q_0 + P^{-1} Q_1 Z_{t-1} + P^{-1} V_t
\]

with

\[
E[Z_t] = (I - P^{-1} Q_1)^{-1} P^{-1} Q_0
\]

and so

\[
E_t [Z_{t+j}] = (I - P^{-1} Q_1)^{-1} P^{-1} Q_0 + (P^{-1} Q_1)^j
\]

\[
\times (Z_t - (I - P^{-1} Q_1)^{-1} P^{-1} Q_0)
\]

\[
= \left( (I - P^{-1} Q_1)^{j} \right) \left( (I - P^{-1} Q_1)^{-1} P^{-1} Q_0 \right) + (P^{-1} Q_1)^j Z_t
\]

and so

\[
E_t [\theta_{t+j}] = g_{0}^{(j)} + \sum_{i=1}^{p} g_{i}^{(j)} \theta_{t+i-1}
\]

where \(g_{0}^{(j)}\) is the first element of \((I - (P^{-1} Q_1)^j) \left( (I - P^{-1} Q_1)^{-1} P^{-1} Q_0 \right)\), and \(g_{i}^{(j)}\) is the \((1, i)\) element of \((P^{-1} Q_1)^j\). Next I use this result to obtain:

\[
E \left[ \Delta C(\mathbf{X}_t) E_t [\theta_{t+j}] \right] = g_{0}^{(j)} E [\Delta C(\mathbf{X}_t)] + g_{1}^{(j)} E [\Delta C(\mathbf{X}_t) \theta_t] + \sum_{i=2}^{p} g_{i}^{(j)} E [\Delta C(\mathbf{X}_t) \theta_{t+i-1}]
\]

so

\[
\begin{aligned}
E \left[ \Delta C(\mathbf{X}_t) \theta_t \right] &= \frac{1}{g_{0}^{(j)}} E \left[ \Delta C(\mathbf{X}_t) E_t [\theta_{t+j}] - g_{0}^{(j)} E \left[ \Delta C(\mathbf{X}_t) \right] \right] \\
&- \sum_{i=2}^{p} g_{i}^{(j)} E [\Delta C(\mathbf{X}_t) \theta_{t+i-1}]
\end{aligned}
\]
which yields
\[
E \left[ \Delta C \left( X_1 \right) \left( \hat{\theta}_{i+j} - \theta_i \right) \right] \\
= \left( 1 - \frac{1}{g_1^i} \right) E \left[ \Delta C \left( X_1 \right) \hat{\theta}_{i+j} \right] + \sum_{i=2}^{p} \frac{g_i^j}{g_1^i} E \left[ \Delta C \left( X_1 \right) \hat{\theta}_{i-1} \right].
\]

since \( E \left[ \Delta C \left( X_1 \right) \theta_{i+1-1} \right] = E \left[ \Delta C \left( X_1 \right) \hat{\theta}_{i+1-1} \right] \) for \( i \geq 2 \) under Assumption R1. With this result we can now compute \( \beta \):
\[
\beta = \sum_{j=1}^{p} \omega_j E \left[ \Delta C \left( X_1 \right) \left( E_i \left[ \hat{\theta}_{i+j} \right] - \theta_i \right) \right] \\
= \sum_{j=1}^{p} \omega_j \left( 1 - \frac{1}{g_1^j} \right) E \left[ \Delta C \left( X_1 \right) \hat{\theta}_{i+j} \right] \\
+ \sum_{j=1}^{p} \omega_j \sum_{i=2}^{p} \frac{g_i^j}{g_1^i} E \left[ \Delta C \left( X_1 \right) \hat{\theta}_{i-1} \right].
\]

(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, \( \hat{\beta}_1 \). This collection of elements is defined as:
\[
\Phi_i = \left[ \Delta (\theta_i, \boldsymbol{X}_j) \right], \quad \Delta (\theta_i, \boldsymbol{X}_j) = \Delta (\theta_i, \boldsymbol{X}_j) \left( \hat{\theta}_{i+1}, \ldots, \hat{\theta}_{i+j} \right),
\]
\[
\Delta (\theta_i, \boldsymbol{X}_j) \hat{\theta}_{i-1}, \ldots, \Delta (\theta_i, \boldsymbol{X}_j) \hat{\theta}_{i-p+1}, \hat{\theta}_i, \hat{\theta}_{i+1}, \ldots, \hat{\theta}_{i+2p}.
\]

with Assumptions A1 and A2 applied to \( \Phi_i \), we have \( \sqrt{T} (\Phi_i - E [\Phi_i]) \rightarrow^d N(0, \sigma^2) \) using Theorem 3 of Politis and Romano (1994).

Note that the last \( 2p + 1 \) elements of \( \Phi_i \) are sufficient to obtain estimates of the mean and the first \( 2p \) autocovariances of \( \hat{\theta}_i \), since \( E \left[ \hat{\theta}_i \right] = E [\theta_i] \) by Assumption P1, and \( E \left[ \hat{\theta}_i \hat{\theta}_{i+j} \right] = E \left[ \left[ \theta_i + \nu_i \right] \left[ \theta_{i+j} + \nu_{i+j} \right] \right] = E \left[ \theta_i \theta_{i+j} \right] \) by Assumptions P1 and T2.

Let \( \gamma_j \equiv \text{Cov} \left[ \theta_i, \theta_{i+j} \right] \), then by the properties of an AR(p) process we have \( \Psi \Phi = \Psi \), where
\[
\Psi = \begin{bmatrix}
\gamma_0 & \gamma_{-1} & \cdots & \gamma_1 \\
\gamma_{p+1} & \gamma_p & \cdots & \gamma_2 \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_{2p-1} & \gamma_{2p-2} & \cdots & \gamma_p \\
\end{bmatrix},
\Psi = \left[ \gamma_0, \gamma_{-1}, \ldots, \gamma_{2p-1} \right], \quad \Phi = \left[ \phi_1, \ldots, \phi_p \right].
\]

and by Assumption T2 we can obtain \( \Phi = \hat{\Psi} - \hat{\Phi} \), where \( \hat{\Psi} \) and \( \hat{\Phi} \) are the equivalents of \( \Psi \) and \( \Phi \) using sample autocovariances rather than population autocovariances. From \( \hat{\Phi} \) we can obtain estimates of \( \hat{P}, \hat{Q}_0, \) and \( \hat{Q}_1 \) and thus estimates of the parameters \( g_i^j \), for \( i = 0, 1, \ldots, p \) and \( j = 1, 2, \ldots, p \), from these we can compute the estimated bias term \( \hat{\beta}_i \). Given asymptotic normality of \( \hat{\beta}_i \) and the fact that \( \frac{1}{T} \sum_{t=1}^{T} \Delta L (Y_t, \boldsymbol{X}_t) - \hat{\beta}_i \) is a smooth function of the elements of \( \hat{\beta}_i \), we can then apply the delta method, see Lemma 2.5 of Hayashi (2000) for example, to obtain asymptotic normality of \( \frac{1}{T} \sum_{t=1}^{T} \Delta L (Y_t, \boldsymbol{X}_t) - \hat{\beta}_i \) 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 \( L (Y_t, \boldsymbol{X}_t) \) given in Eq. (5) around \( (\theta_i, \boldsymbol{X}_i) \):
\[
L (Y_t, \boldsymbol{X}_t) = L (\theta_t, \boldsymbol{X}_t) + \frac{dL (\theta_t, \boldsymbol{X}_t)}{d\theta} (Y_t - \theta_t) \]
\[
= \frac{1}{2} \frac{\partial^2 L (\theta, \boldsymbol{X})}{\partial \theta^2} (Y_t - \theta_t)^2 + \frac{1}{2} \frac{\partial^2 L (\theta, \boldsymbol{X})}{\partial \theta^2} (Y_t - \theta_t)^2,
\]
where \( \theta_t = \lambda_t \theta_t + (1 - \lambda_t) Y_t \) for some \( \lambda_t \in [0, 1] \), and using the functional form of \( L \) in Eq. (5). Thus
\[
\Delta L (Y_t, \boldsymbol{X}_t) = \Delta L (\theta_t, \boldsymbol{X}_t) + \Delta C (\theta_t, \boldsymbol{X}_t) (Y_t - \theta_t)
\]
where
\[
\Delta C (\theta_t, \boldsymbol{X}_t) = \begin{bmatrix}
C (X_{1t}) - C (X_{2t}) \\
\vdots \\
C (X_{ht}) - C (X_{ht})
\end{bmatrix}.
\]

Next, note:
\[
E \left[ \Delta C \left( X_1 \right) (Y_t - \theta_t) \right] \\
= \left[ \frac{\partial^2 L (\theta, \boldsymbol{X})}{\partial \theta^2} (Y_t - \theta_t)^2 + \frac{1}{2} \frac{\partial^2 L (\theta, \boldsymbol{X})}{\partial \theta^2} (Y_t - \theta_t)^2 \right] \left( \sum_{i=1}^{p} \omega_i E \left[ \nu_{i+1} \right] \right).
\]

by the law of iterated expectations, since \( \nu_{i+1} \subset \mathcal{F}_t \). This thus yields \( E \left[ \Delta L (\theta_t, \boldsymbol{X}_t) \nu_{i+1} \right] = E \left[ \Delta L (Y_t, \boldsymbol{X}_t) \nu_{i+1} \right] \) as claimed.

(b) Using Exercise 5.21 of White (2001) for example, we have \( \sqrt{T} (\hat{\alpha} - \alpha) \rightarrow^d N(0, I) \), where \( \hat{\alpha} \) is given in the statement of the proposition.

To show that \( \alpha = \alpha \) note that \( \Delta L (Y_t, \boldsymbol{X}_t) = \Delta L (\theta_t, \boldsymbol{X}_t) + \Delta C (\theta_t) (Y_t - \theta_t) = \alpha' Z_{-1} + e_t + \Delta C (\theta_t) (Y_t - \theta_t) \) \( \equiv \alpha' Z_{-1} + e_t \), with \( E \left[ e_t \right] = E \left[ e_t Z_{-1} \right] + E \left[ \Delta C (\theta_t) (Y_t - \theta_t) Z_{-1} \right] = 0 \), since \( E \left[ \Delta C (\theta_t) (Y_t - \theta_t) Z_{-1} \right] = E \left[ E \left[ \Delta C (\theta_t) (Y_t - \theta_t) \right] Z_{-1} \right] = 0 \) by (c), and \( E \left[ e_t Z_{-1} \right] = 0 \) under A3. Thus \( \alpha = \alpha \) as claimed. □

**Proof of Proposition 5.** (a) We first obtain \( E \left[ \Delta L (Y_t, \boldsymbol{X}_t) Z_{-p} \right] \) using calculations previously presented in the proof of Proposition 3.
\[ E \left[ \Delta L(Y_t, X_t) Z_{t-p} \right] = E \left[ \Delta L(\theta_t, X_t) Z_{t-p} \right] \]

\[ = \sum_{j=1}^{p} \omega_j E \left[ \Delta C(X_t) \left( \hat{\theta}_{t+j} - \theta_t \right) Z_{t-p} \right] \]

\[ E \left[ \Delta C(X_t) \left( \hat{\theta}_{t+j} - \theta_t \right) Z_{t-p} \right] \]

\[ = g^{ij}_0 E \left[ \Delta C(X_t) Z_{t-p} \right] + \sum_{i=2}^{p} g^{ij}_i E \left[ \Delta C(X_t) \hat{\theta}_{t+i-1}, Z_{t-p} \right] \]

\[ + \left( g^{ij}_1 - 1 \right) E \left[ \Delta C(X_t) \theta_t Z_{t-p} \right] \]

and

\[ E \left[ \Delta C(X_t) \theta_t Z_{t-p} \right] = \frac{1}{\phi_0} E \left[ \Delta C(X_t) \hat{\theta}_{t+1} Z_{t-p} \right] \]

\[ - \frac{\phi_0}{\phi_1} E \left[ \Delta C(X_t) \hat{\theta}_{t+1}, Z_{t-p} \right] \]

\[ - \sum_{j=2}^{p} \frac{\phi_j}{\phi_1} E \left[ \Delta C(X_t) \hat{\theta}_{t+1}, Z_{t-p} \right]. \]

Pulling these results together we obtain:

\[ E \left[ \Delta L(Y_t, X_t) Z_{t-p} \right] = E \left[ \Delta L(\theta_t, X_t) Z_{t-p} \right] \]

\[ = \sum_{j=1}^{p} \omega_j E \left[ \Delta C(X_t) \left( \hat{\theta}_{t+j} - \theta_t \right) Z_{t-p} \right] \]

\[ = \sum_{j=1}^{p} \omega_j \left\{ g^{ij}_0 E \left[ \Delta C(X_t) Z_{t-p} \right] \right\} \]

\[ + \sum_{i=2}^{p} g^{ij}_i E \left[ \Delta C(X_t) \hat{\theta}_{t+i-1}, Z_{t-p} \right] \]

\[ + \left( g^{ij}_1 - 1 \right) \left\{ \frac{1}{\phi_1} E \left[ \Delta C(X_t) \hat{\theta}_{t+1}, Z_{t-p} \right] - \frac{\phi_0}{\phi_1} E \left[ \Delta C(X_t) \hat{\theta}_{t+1}, Z_{t-p} \right] \right\} \]

\[ - \sum_{j=2}^{p} \frac{\phi_j}{\phi_1} E \left[ \Delta C(X_t) \hat{\theta}_{t+1}, Z_{t-p} \right]. \]

Thus with \( \Delta L(\hat{\theta}_t, X_t) \) defined as in the proposition, we obtain

\[ E \left[ \Delta L(\hat{\theta}_t, X_t) Z_{t-p} \right] = E \left[ \Delta L(\theta_t, X_t) Z_{t-p} \right]. \]

(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:

\[ D_t = \left[ \Delta L(\hat{\theta}_t, X_t) Z_{t-p}, \Delta C(X_t) \hat{\theta}_{t+1}, Z_{t-p}, \ldots, \Delta C(X_t) \hat{\theta}_{t+1}, Z_{t-p} \right] \]

\[ \ldots, \Delta C(X_t) \hat{\theta}_{t+1}, Z_{t-p} \].

and with Assumptions A1 and A2 applied to \( D_t \), we have

\[ \sqrt{T} \left( D_t - E(D_t) \right) \rightarrow^d N \left( 0, V_T \right) \]

using Theorem 3 of Politis and Romano (1994). As in the proof of Proposition 3, the last \( 2p + 1 \) elements of \( D_t \) are sufficient to estimate \( P, Q_0, \) and \( Q_1 \) and thus estimates of the parameters \( g^{ij}_0, \) 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 \( D_t \) and the fact that \( \hat{\theta}_t \) is a smooth function of the elements of \( D_t, \) we can then apply the delta method, see Lemma 2.5 of Hayashi (2000) for example, to show asymptotic normality of \( (\hat{\theta}_t - \theta) \), and obtain its covariance matrix. To show that \( \hat{\theta}_n = \theta \) we use the result from part (a) which provides

\[ \hat{\theta}_n = \left( E \left[ Z_{t-p} Z_{t-p} \right] \right)^{-1} E \left[ Z_{t-p} \Delta L(\hat{\theta}_t, X_t) \right] =
\]

\[ E \left[ Z_{t-p} Z_{t-p} \right]^{-1} E \left[ Z_{t-p} \Delta L(\hat{\theta}_t, X_t) \right] \equiv \alpha. \]

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

References


