Data-Based Ranking of Realised Volatility Estimators*

Andrew J. Patton

University of Oxford

First version: April 2007. This version: 31 March 2009

Abstract

This paper presents new methods for formally comparing the accuracy of estimators of the quadratic variation of a price process. I provide conditions under which the relative average accuracy of competing estimators can be consistently estimated (as $T \to \infty$) from available data, and show that existing tests from the forecast evaluation literature may be adapted to the problem of ranking these estimators. The proposed methods eliminate the need for specific assumptions about the properties of the microstructure noise, and the need to estimate quantities such as integrated quarticity or the noise variance, and facilitate comparisons of estimators that would be difficult using methods from the extant literature, such as those based on different sampling schemes (calendar-time vs. tick-time). In an application to high frequency IBM stock price data between 1996 and 2007, I find that tick-time sampling is generally preferable to calendar-time sampling, and that the optimal sampling frequency is between 15 seconds and 5 minutes, when using standard realised variance.

Keywords: realized variance, volatility forecasting, forecast comparison.

J.E.L. codes: C52, C22, C53.

*I thank 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 SITE workshop in Stanford, and the workshop on Model Evaluation and Predictive Ability in Paris for helpful comments and suggestions. Runquan Chen provided excellent research assistance. Financial support from the Leverhulme Trust under grant F/0004/AF is gratefully acknowledged. Contact address: Department of Economics, and Oxford-Man Institute of Quantitative Finance, University of Oxford, Manor Road, Oxford OX1 3UQ, United Kingdom. Email: andrew.patton@economics.ox.ac.uk.

1 Introduction

The past decade has seen an explosion in research on volatility measurement, as distinct from volatility forecasting[1]. This research has focussed 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). These "realised volatility" (RV) estimators or "realised measures" generally aim at measuring the quadratic variation or integrated variance of the log-price process of some asset or collection of assets.

This profusion of research has lead to a need for some practical guidance on which RV estimator to select for a given empirical analysis. In addition to the particular estimator to use, the performance of RV estimators is generally affected by the frequency used to sample the price process (for example, every 5 minutes or every 30 seconds), see Zhou (1996) and Bandi and Russell (2008) for example, and may also be affected by the decision to sample in calendar time or in "tick time" (for example, every r minutes or every s trades), and the decision to use prices from transactions or from quotes, see Bandi and Russell (2006b), Hansen and Lunde (2006a) and Oomen (2006).

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

$$E\left[\Delta L\left(\theta_t, \mathbf{X}_t\right)\right] \equiv E\left[L\left(\theta_t, X_{i,t}\right)\right] - E\left[L\left(\theta_t, X_{j,t}\right)\right] \tag{1}$$

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

[1] See Andersen and Bollerslev (1998), Andersen, *et al.* (2001a, 2003), Barndorff-Nielsen and Shephard (2002, 2004), Aït-Sahalia, *et al.* (2005), Zhang, *et al.* (2005), Hansen and Lunde (2006a), Christensen and Podolskij (2007), and Barndorff-Nielsen, *et al.* (2008a) amongst many others. Andersen, *et al.* (2006) and Barndorff-Nielsen and Shephard (2007) present recent surveys of this burgeoning field. Older papers on this topic include Merton (1980), French, *et al.* (1987) and Zhou (1996).

With an estimator of $E\left[\Delta L\left(\theta_t, \mathbf{X}_t\right)\right]$ in hand, it is possible to employ one of the many tests from the literature on forecast evaluation and comparison, such as Diebold and Mariano (1995) and West (1996) for pair-wise comparisons, White (2000), Hansen (2005), Hansen, *et al.* (2005) and Romano and Wolf (2005) for comparisons involving a large number of RV estimators, and Giacomini and White (2006) for conditional comparisons of RV estimators. These tests rely on standard large sample asymptotics ($T \rightarrow \infty$) rather than continuous-record asymptotics ($m \rightarrow \infty$), and thus can be used to compare the "finite m" performance of different estimators. I provide conditions under which these tests can be applied to the problem of ranking RV estimators. The proposed methods rely on the existence of a volatility proxy that is unbiased for the latent target variable, θ_t, and satisfies an uncorrelatedness condition, described in detail below. This proxy must be unbiased but it does not need to be very precise; a simple and widely-available proxy is the daily squared return, for example.

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

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

where p_{τ_i} is the log-price at time τ_i, $\{\tau_0, \tau_1, ..., \tau_m\}$ are the times at which the price of the asset is available during period t, and m is the number of intra-period observations used in computing the estimator. In the absence of market microstructure effects, distribution theory for the simplest RV estimator would suggest sampling prices as often as possible, see Merton (1980) 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[2], and several papers have attempted to address this problem: Zhou (1996) derives the mean-squared-error (MSE) optimal sampling frequency (or, equivalently, optimal choice

[2] Early research in this area, see Zhou (1996) and Andersen, *et al.* (2000), employed "volatility signature plots" to show graphically that sampling at the highest possible frequency is not optimal in practice: at very high frequencies, effects such as bid-ask bounce and stale prices can lead to large biases in simple RV estimators. More sophisticated estimators, such as the two-scale estimator of Zhang, *et al.* (2006) and the realised kernel estimator of Barndorff-Nielsen, *et al.* (2008a) provide consistent estimates of quadratic variation using very high frequency data, under some conditions, by taking these autocorrelations into account in the construction of the estimator.

of m) for the $RVAC1$ estimator assuming *iid* noise and intra-daily homoskedasticity[3]; Aït-Sahalia, *et al.* (2005) derive the MSE-optimal choice of m for the standard RV estimator under a variety of cases (*iid* noise, serially correlated noise, and noise correlated with the efficient price); Andersen, *et al.* (2007) derive the MSE-optimal choice of m for the $RVACq$ estimator, the realised kernel estimator of Barndorff-Nielsen, *et al.* (2008a) and the two-scale estimator of Zhang, *et al.* (2006), under the assumption of *iid* noise; Hansen and Lunde (2006a) derive the MSE-optimal choice of m for $RVACq$ estimators assuming *iid* noise; Bandi and Russell (2006a,c) derive the optimal choice of the m for standard RV, and the optimal *ratio* of q/m for $RVACq$ estimators using m intra-daily observations, under the assumption of *iid* noise; Bandi, *et al.* (2007) consider the optimal choice of m when the noise process is conditionally mean zero but potentially heteroskedastic; and Barndorff-Nielsen, *et al.* (2008a) examine the optimal sampling frequency and number of lags to use with a variety of realised kernel estimators, under the assumption of *iid* noise. While the methods of these papers differ, they have in common their use of continuous-record asymptotics in their derivations, the use of 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, c.f. Hansen and Lunde (2006a), Kalnina and Linton (2007), 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.* (2007) 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, which are not required in most of the existing literature, and so the proposed tests complement, rather than substitute, existing methods; they provide an alternate approach to

[3] The $RVACq$ estimator adjusts the standard RV estimator in equation (2) to account for autocovariances up to order q.

[4] Gatheral and Oomen (2007) provide an alternative analysis of the problem of choosing an RV estimator via a detailed simulation study.

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

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 (2005) 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 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. I find that Romano-Wolf (2005) tests clearly reject the squared daily return in favour of a 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-minute calendar-time RV: estimators based on data sampled at between 15 seconds and 2 minutes are significantly more accurate than 5-minute RV. In general, I find that using tick-time sampling leads to more accurate RV estimation than using calendar-time sampling, particularly when trades 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 of the data.

The remainder of the paper is structured as follows. Section 2 presents the main theoretical results of this paper, and discusses the important differences between comparing estimators of quadratic variation and comparing volatility forecasts. Section 3 presents a simulation study of the methods of this paper for a realistic stochastic volatility process, and Section 4 presents an application using high frequency quote and trade data on IBM over the period January 1996 to June 2007. Section 5 concludes, and all proofs are collected in the Appendix.

2 Data-based ranking of RV estimators

2.1 Notation and background

The target variable, generally quadratic variation (QV) or integrated variance[6] (IV), is denoted θ_t. I assume that θ_t is $\mathcal{F}_t$-measurable, where $\mathcal{F}_t$ is the information set generated by the complete path of the log-price process. For the remainder of the paper I assume that θ_t is a scalar; I discuss the extension to vector (or matrix) target variables in the conclusion. The estimators of θ_t are denoted $X_{i,t}$, $i = 1, 2, ..., k$. Often these will be the same estimator applied to data sampled at different frequencies, for example 1-minute returns vs. 30-minute 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, $X_{i,t}$, to the target variable, θ_t. Two popular (pseudo-)distance measures in the volatility literature are MSE and QLIKE:

$$\begin{aligned} \text{MSE} \quad L(\theta, X) &= (\theta - X)^2 && (3) \\ \text{QLIKE} \quad L(\theta, X) &= \frac{\theta}{X} - \log\left(\frac{\theta}{X}\right) - 1 && (4) \end{aligned}$$

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 (2008), which nests MSE and QLIKE as special cases:

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

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

For the remainder of the paper I will use the following notation to describe the ($k-1$ vector

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

of) differences in the distances from the target variable to a collection of RV estimators:

$$\Delta L(\cdot, \mathbf{X}_t) \equiv \begin{bmatrix} L(\cdot, X_{1t}) - L(\cdot, X_{2t}) \\ \vdots \\ L(\cdot, X_{1t}) - L(\cdot, X_{kt}) \end{bmatrix} \qquad (6)$$

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

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

2.2 Ranking volatility forecasts vs. 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 (2008) 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.

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

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

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

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

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

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

2.3 Ranking RV estimators

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

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

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

In the propositions below I consider using a convex function of leads of $\tilde{\theta}_t$.

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

Using leads of the proxy is important for breaking the correlated measurement errors problem, which makes it possible to overcome the problems identified in Proposition 1. Y_t is thus interpretable as an instrument for $\tilde{\theta}_t$. Our focus on differences in average accuracy makes this a *non-linear* instrumental variables problem, and like other such problems it is not sufficient to simply assume that $Corr\left[Y_t, \tilde{\theta}_t\right] \neq 0$; some more structure is required. I obtain results in this application by considering two alternative approximations of the conditional mean of θ_t.

Numerous papers on the conditional variance (see Bollerslev, *et al.*, 1994, Engle and Patton, 2001, and Andersen, *et al.*, 2006 for example), or integrated variance (see Andersen, *et al.*, 2004, 2007) have reported that these quantities are very persistent, close to being (heteroskedastic) random walks. The popular RiskMetrics model, for example, is based on a unit root assumption for the conditional variance. It should be noted that Wright (1999) provides thorough evidence *against* the presence of a unit root in daily conditional variance for several assets, however, despite this, it has proven to be a good approximation in many cases. Given this, consider the following assumption:

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

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 equation (5). Then*

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

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

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

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

where Ω_1 is given in the proof.

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

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

Part (a) of the above proposition shows that it is possible to obtain an unbiased estimate of the difference in the average distance from the latent target variable, θ_t, using a suitably-chosen volatility proxy, under certain conditions. This opens the possibility to use existing methods from the forecast evaluation literature to help us choose between RV estimators. 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-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, *et al.* (2005), 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., $T \to \infty$ rather than $m \to \infty$), on mixing and moment conditions, and on the applicability of the random walk approximation for the latent target variable. In Section 3 below I show that these assumptions are reasonable in a realistic simulation design.

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

Assumption T2: $\theta_t = \phi_0 + \sum_{i=1}^{p} \phi_i \theta_{t-i} + \eta_t$, *with* $E[\eta_t | \mathcal{F}_{t-1}] = 0$, $\phi_1 \neq 0$, *the matrix* Ψ *defined in equation (35) is invertible, and* $\boldsymbol{\phi} \equiv [\phi_1, ..., \phi_p]'$ *is such that* θ_t *is covariance stationary.*

When the order of the autoregression is greater than one, I also require assumption R1, below. This assumption is plausible for most RV estimators in the literature, as they are generally based on data from a single day, although Barndorff-Nielsen, *et al.* (2004) and Owens and Steigerwald (2007) are two exceptions.

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

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

$$\begin{aligned}
(a) \qquad E\left[\Delta L\left(\theta_t, \mathbf{X}_t\right)\right] &= E\left[\Delta L\left(Y_t, \mathbf{X}_t\right)\right] - \boldsymbol{\beta} \\
\text{where } \boldsymbol{\beta} &= E\left[\Delta C\left(\mathbf{X}_t\right)\right] \sum_{j=1}^{J} \omega_j \frac{g_0^{(j)}}{g_1^{(j)}} + \sum_{j=1}^{J} \omega_j \left(1 - \frac{1}{g_1^{(j)}}\right) E\left[\Delta C\left(\mathbf{X}_t\right) \tilde{\theta}_{t+j}\right] \\
&\quad + \sum_{j=1}^{J} \omega_j \sum_{i=2}^{p} \frac{g_i^{(j)}}{g_1^{(j)}} E\left[\Delta C\left(\mathbf{X}_t\right) \tilde{\theta}_{t+1-i}\right]
\end{aligned}$$

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

(b) If we further assume A1 and A2 in the Appendix hold for the series $\mathbf{B}_t$, defined in equation (34), then:

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

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

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

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

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

Proposition 3 relaxes the assumption of a random walk, at the cost of introducing a bias term to the expected loss computed using the proxy. This bias term, however, can be consistently estimated under the assumption that the target variable follows a stationary AR(p) process. The cost of the added flexibility in allowing for a general AR(p) process for the target variable is the added estimation error induced by having to estimate the AR(p) parameters, and having to estimate additional terms of the form $E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+j}\right]$. This estimation error will lead to reduced power to distinguish between competing RV estimators than would otherwise be the case.

2.3.2 Conditional rankings of RV estimators

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

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

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

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

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

$$\begin{aligned} H_0 &: \boldsymbol{\alpha} = 0 \\ \text{vs. } H_a &: \boldsymbol{\alpha} \neq 0 \end{aligned} \tag{9}$$

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

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

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

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

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

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

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

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

$$\text{where} \quad \hat{D}_T \equiv \hat{M}_T^{-1}\hat{\Omega}_T\hat{M}_T^{-1}, \;\; \hat{M}_T \equiv \frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_{t-1}\mathbf{Z}_{t-1}', \;\; \Omega_T \equiv V\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{Z}_{t-1}e_t\right]$$

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

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

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

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

while the adjusted regression becomes:

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

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

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

This adjusted dependent variable is constructed such that $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha}$, and thus estimating equation (12) by OLS yields a consistent estimator of the unknown true parameter $\boldsymbol{\alpha}$. (Note that if $\phi_0 = 0$

and $\phi_1 = 1$, which corresponds to the random walk case, the adjustment term drops out and we obtain the same result as in Proposition 4.) Of course, the parameters of the AR(p) process must be estimated, leading to a feasible adjusted regression:

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

where

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

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

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

Assumption R1': *X_t is conditionally independent of ν_{t-j} given $\mathcal{F}_{t-j-1}$, for all $j > 0$.*

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

$$\begin{aligned}
\Delta\widetilde{L(\theta_t, \mathbf{X}_t)} &\equiv \Delta L(Y_t, \mathbf{X}_t) + \lambda_0\Delta C(\mathbf{X}_t) + \lambda_1\Delta C(\mathbf{X}_t)\,\tilde{\theta}_{t+1} + \sum_{i=2}^{p}\lambda_i\Delta C(\mathbf{X}_t)\,\tilde{\theta}_{t+1-i} \\
\text{where}\quad \lambda_1 &= \frac{1}{\phi_1} - \sum_{j=1}^{J}\omega_j\frac{g_1^{(j)}}{\phi_1},\ \text{and}\ \lambda_i = -\frac{\phi_i}{\phi_1} - \sum_{j=1}^{J}\omega_j\left(g_i^{(j)} - g_1^{(j)}\frac{\phi_i}{\phi_1}\right),\quad i = 0, 2, 3, ..., p \\
\text{and}\quad \Delta\widehat{L(\theta_t, \mathbf{X}_t)} &\equiv \Delta L(Y_t, \mathbf{X}_t) + \hat{\lambda}_{0,T}\Delta C(\mathbf{X}_t) + \hat{\lambda}_{1,T}\Delta C(\mathbf{X}_t)\,\tilde{\theta}_{t+1} + \sum_{i=2}^{p}\hat{\lambda}_{i,T}\Delta C(\mathbf{X}_t)\,\tilde{\theta}_{t+1-i}
\end{aligned}$$

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

(a)
$$E\left[\Delta\widetilde{L(\theta_t, \mathbf{X}_t)}\mathbf{Z}_{t-p}\right] = E\left[\Delta L(\theta_t, \mathbf{X}_t)\,\mathbf{Z}_{t-p}\right] \quad \text{for any } \mathbf{Z}_{t-p} \in \mathcal{F}_{t-p}$$

(b) Denote the OLS parameter estimate of $\tilde{\tilde{\boldsymbol{\alpha}}}$ in equation (14) as $\hat{\boldsymbol{\alpha}}_T$. If we further assume A1 and A2 in the Appendix hold for the series $\mathbf{D}_t$, defined in equation (36), then:

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

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

$$\sup_z \left|P^* \left[\|\hat{\boldsymbol{\alpha}}_T^* - \hat{\boldsymbol{\alpha}}_T\| \leq z\right] - P\left[\|\hat{\boldsymbol{\alpha}}_T - \boldsymbol{\alpha}\| \leq z\right]\right| \to^p 0, \quad \text{as } T \to \infty.$$

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

3 Simulation study

To examine the finite-sample performance of the results in the previous section, I present the results of a small simulation study. I use a log-normal stochastic volatility model with a leverage effect, with the same parameters as in Gonçalves and Meddahi (2009):

$$\begin{aligned} d\log P^*(t) &= 0.0314d(t) + \nu(t)\left(-0.576dW_1(t) + \sqrt{1-0.576^2}dW_2(t)\right) \\ d\log\nu^2(t) &= -0.0136\left(0.8382 + \log\nu^2(t)\right)d(t) + 0.1148dW_1(t) \end{aligned} \tag{16}$$

In simulating from these processes I use a simple Euler discretization scheme, with the step size calibrated to one second (i.e., with 23,400 steps per simulated trade day, which assumed to be 6.5 hours 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 *iid* error term for the observed log-price:

$$\begin{aligned} \log P(t_j) &= \log P^*(t_j) + \xi(t_j) \\ \xi(t_j) &\sim iid\ N\left(0, \sigma_\xi^2\right) \end{aligned} \tag{17}$$

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

$$\frac{2\sigma_\xi^2}{V[r_t]\frac{5}{390} + 2\sigma_\xi^2} = 0.20 \tag{18}$$

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 process to be simulated above exhibits a leverage effect and is 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 the realistic case that 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 is investigated via the following experiment. For simplicity I focus on pair-wise comparisons of RV estimators, each implemented using the 1000 draws from the stationary bootstrap of Politis and Romano (1994), thus making this a 'reality check'-type test from White (2000). I set the each RV estimator equal to the true QV plus some noise:

$$\begin{aligned} X_{it} &= QV_t + \zeta_{it}, \;\; i = 1,2 && (19) \\ \zeta_{1t} &= \omega \nu_t^{30\,\min} + (1-\omega)\,\sigma_u U_{1t} && (20) \\ \zeta_{2t} &= \omega \nu_t^{30\,\min} + (1-\omega)\,\sigma_u U_{2t} + \sqrt{\sigma_{\zeta 2}^2 - \sigma_{\zeta 1}^2}\,U_{3t} && (21) \\ [U_{1t}, U_{2t}, U_{3t}]' &\sim iid\ N(0, I) \\ \text{where }\ \nu_t^{30\,\min} &\equiv RV_t^{30\,\min} - IV_t \end{aligned}$$

The above structure allows the measurement error on each of the RV estimators to be correlated with the proxy measurement error, consistent with what is faced in practice. As a benchmark, I use the measurement errors on $RV^{30\,\min}$ to generate this correlation, and I set the correlation to be $\rho = Corr\left[\nu_t^{30\,\min}, \zeta_{1t}\right] = 0.5$, by setting the parameters $\left(\omega, \sigma_u^2\right)$ using equations (22) and (23) below. The equations below also allow me to vary the variance of the errors associated with the RV estimators, $\sigma_{\zeta 1}^2$ and $\sigma_{\zeta 2}^2$. In the study of the size of the tests I set $\sigma_{\zeta 1}^2 / V\left[IV_t\right] = \sigma_{\zeta 2}^2 / V\left[QV_t\right] = 0.1$ (and so the variable U_{3t} drops out of equation 21) which is approximately equal to $V\left[\ \nu_t^{30\,\min}\right] / V\left[QV_t\right]$ in this simulation. To study the power, I fix $\sigma_{\zeta 1}^2 / V\left[QV_t\right] = 0.1$, and let $\sigma_{\zeta 2}^2 / V\left[QV_t\right] =$ 0.15, 0.2, 0.5, 1.

$$\begin{aligned} \omega &= \frac{\rho \sigma_{\zeta 1}}{\sigma_\nu} && (22) \\ \sigma_u^2 &= \frac{\sigma_\nu^2 \sigma_{\zeta 1}^2 \left(1-\rho^2\right)}{\left(\sigma_\nu - \rho\sigma_{\zeta 1}\right)^2} && (23) \\ \text{where }\ V\left[\nu_t^{30\,\min}\right] &\equiv \sigma_\nu^2 \end{aligned}$$

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-minute 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 MSE and QLIKE pseudo-distance measures from equations (3) and (4).

The first row of each panel of Table 1 corresponds to the case when the null hypothesis is satisfied, and thus we expect these figures to be close to 0.05, the nominal size of the tests. For both MSE and QLIKE and for $T = 500$ and $T = 2500$ we see that the finite-sample size is reasonable, with rejection frequencies reasonably close to 0.05. Most tests appear to be under-sized, meaning that they are conservative tests of the null. The results for the power of the tests are as expected: the power of the new tests are worse than would be obtained if the true QV were observable; power is greater when using a longer time series of data; power is worse when a noisier instrument is used (true QV vs. 30-minute RV vs. daily squared returns); and the power of the test based on the AR(1) approximation is worse than that based on the random walk approximation. The AR(1) approximation has little power when the volatility proxy is very noisy and T is small: in that case it appears that the estimation of the AR parameters overwhelms any information about the relative accuracy of the two RV estimators. In this particular design, the power of the tests based on the MSE loss function is greater than those based on QLIKE loss, though this is likely due to the additive nature of the noise in the design of the RV estimators being compared.

[INSERT TABLE 1 ABOUT HERE]

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

$$X_{1t} = QV_t + \zeta_{1t} \tag{24}$$

$$X_{2t} = QV_t - \lambda QV_{t-1} + \zeta_{2t} \tag{25}$$

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

$$[U_{1t}, U_{2t}]' \sim iid\ N(0, I)$$

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

$$L\left(\tilde{\theta}_{t+1}, X_{1t}\right) - L\left(\tilde{\theta}_{t+1}, X_{2t}\right) = \alpha_0^u + e_t^u, \quad \text{or} \tag{26}$$

$$L\left(\tilde{\theta}_{t+1}, X_{1t}\right) - L\left(\tilde{\theta}_{t+1}, X_{2t}\right) = \alpha_0 + \alpha_1 \log \frac{1}{10} \sum_{j=1}^{10} \tilde{\theta}_{t-j} + e_t \tag{27}$$

where $\tilde{\theta}_t$ is the volatility proxy: daily squared returns, 30-minute RV or the true QV. I use Propositions 4 and 5 to consider three tests based on the above regressions: an unconditional test $(\alpha_0^u = 0)$, a test that the slope coefficient in the second regression is zero $(\alpha_1 = 0)$, or a joint conditional test $(\alpha_0 = \alpha_1 = 1)$. 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 available on request[7].

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 of the null is lower when a less accurate volatility proxy is employed, higher when a long time series of data is available, and higher using the random walk approximation than using the AR(1) approximation.

[INSERT TABLES 2 AND 3 ABOUT HERE]

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 (9:45am to 4pm) continuously-compounded return on IBM. I use data on NYSE

[7] Like those using QLIKE, the tests using MSE distance are slightly under-sized, though less so than using QLIKE. Related to this, the power of tests using MSE in this simulation are slightly higher than those using QLIKE. I focus on QLIKE in this section as this is the distance measure used in the empirical work presented in Section 4.

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: the minimum tick size moved from one-eighth of a dollar to one-sixteenth of a dollar on June 24, 1997, and to pennies on January 29, 2001[8]. Further, volatility for this stock (and for the market generally) was high over the early and middle parts of the sample, and very low, by historical standards, in the later years of the sample, see Figure 1. These changes motivate the use of sub-samples in the empirical analyses below: I break the sample into three periods (1996-1999, 2000-2003 and 2004-2007) to determine whether these changes impact the ranking of the competing realised volatility estimators.

I consider standard realised variance, as presented in equation (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[9] 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-second sampling and 1-day sampling) which brings the number of RV estimators to 48 in total. In Figure 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.

[INSERT FIGURES 1 AND 2 ABOUT HERE]

In Figures 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

[8] Source: New York Stock Exchange web site, http://www.nyse.com/about/history/timeline_chronology_index.html.

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

variation of the IBM price process, relative to the corresponding distance using 5-minute calendar-time RV on trade prices[10], using the QLIKE distance measure presented in equations (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-minute calendar-time RV on trade prices as the volatility proxy to compute the differences in average distances[11]. I present these results for the full sample and for three sub-samples (1996-1999, 2000-2003, 2004-2007).

The conclusion from these pictures is that there are clear gains to using intra-daily data to compute RV, consistent with the voluminous literature to date: the estimated average distances to the true QV for estimators based on returns sampled at 30-minute 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 seconds or 1 minute, and the best-performing estimator over the full sample is RV based on trade prices sampled in tick time at 1-minute average intervals. The AR approximation gives the same result for the full sample and similar results in the sub-samples.

[INSERT FIGURES 3 AND 4 ABOUT HERE]

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:

$$\begin{aligned} H_0^{(i)} &: E\left[L\left(\theta_t, X_{t,0}\right)\right] = E\left[L\left(\theta_t, X_{t,s}\right)\right], \text{ for } i = 1, 2, ..., 47 \\ \text{vs. } H_1^{(i)} &: E\left[L\left(\theta_t, X_{t,0}\right)\right] > E\left[L\left(\theta_t, X_{t,s}\right)\right] \\ \text{or } H_2^{(i)} &: E\left[L\left(\theta_t, X_{t,0}\right)\right] < E\left[L\left(\theta_t, X_{t,s}\right)\right] \end{aligned}$$

[10] 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.

[11] Using the assumption that the squared open-to-close return is unbiased for the true quadratic variation, I tested whether 5-minute 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.

and identifies which individual null hypotheses, $H_0^{(i)}$, 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 a RV estimator based on 5-minute 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-minute 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-second to 2-minute sampling frequencies are significantly better than 5-minute RV. Estimators with even higher sampling frequencies are not significantly different, while estimators based on 15-minute or lower sampling are found to be significantly worse. The results also indicate that trade prices are preferred to quote prices for most of this sample period. Only in the last sub-sample are quote prices at 15-second to 2-minute sampling frequencies found to out-perform 5-minute 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-minute RV estimator using the Romano-Wolf method, suggesting that the gains from moving beyond 5-minute sampling are hard to identify in the presence of additional estimation error from the AR model, consistent with the simulation results in Section 3.

[INSERT TABLES 4 AND 5 ABOUT HERE]

4.2 Conditional comparisons of RV estimators

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

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

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

$$L\left(\theta_t, X_{t,0}\right) - L\left(\theta_t, X_{t,s}\right) = \beta_0 + \beta_1 Z_{t-1} + e_t \tag{28}$$

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

$$\begin{aligned} H_0 &: \beta_0 = \beta_1 = 0 \\ \text{vs. } H_a &: \beta_i \neq 0 \text{ for some } i = 0, 1 \end{aligned} \tag{29}$$

4.2.1 High-frequency vs. 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 (2004) 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 estimated the following regression, and obtained the results below, with robust t-statistics presented in parentheses below the parameter estimates:

$$L\left(Y_t, RV_t^{(daily)}\right) - L\left(Y_t, RV_t^{(5\,\text{min})}\right) = \underset{(8.42)}{33.67} + e_t \tag{30}$$

$$L\left(Y_t, RV_t^{(daily)}\right) - L\left(Y_t, RV_t^{(5\,\text{min})}\right) = \underset{(11.10)}{24.94} + \underset{(2.55)}{17.85} Z_{t-1} + e_t \tag{31}$$

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

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

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

$$L\left(Y_t, RV_t^{(daily)}\right) - L\left(Y_t, RV_t^{(5\,\text{min})}\right) = \underset{(6.69)}{33.54} + e_t \tag{32}$$

$$L\left(Y_t, RV_t^{(daily)}\right) - L\left(Y_t, RV_t^{(5\,\text{min})}\right) = \underset{(5.75)}{19.93} + \underset{(3.45)}{27.76} Z_{t-1} + e_t \tag{33}$$

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

4.2.2 Tick-time vs. calendar-time sampling

I next use the GW test of conditional accuracy to compare calendar-time sampling with tick-time sampling. Theoretical comparisons of tick-time and calendar-time sampling requires assumptions on the arrival rate of trades, while the methods presented in this paper allow us to avoid making any specific assumptions about the trade arrival process. For example, in a parametric "pure jump" model of high frequency asset prices, Oomen (2006) finds that tick-time sampling leads to more accurate RV estimators than calendar-time sampling when trades arrive at irregular intervals. In general, if the trade arrival rate is correlated with the level of volatility, consistent with the work of Easley and O'Hara (1992), Engle (2000) and Manganelli (2005), then using tick-time sampling serves to make the sampled high-frequency returns closer to homoskedastic, which theoretically should improve the accuracy of RV estimation, see Hansen and Lunde (2006a) and Oomen (2006). I use the log volatility of trade durations to measure how irregularly-spaced trade observations are: this volatility will be zero if trades arrive at evenly-spaced intervals, and increases as trades arrive more irregularly.

I estimate a regression of the difference in the accuracy of a calendar-time RV estimator and a tick-time estimator with the same average sampling frequency, on a constant and the lagged log volatility of trade durations, for each of the frequencies considered in the earlier sections[12], and present the results in Table 6. The first column of Table 6 reports a Diebold and Mariano (1995)-type test of the difference in unconditional average accuracy, across the sampling frequencies, using

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

the RW approximation. This difference is positive and significant for the highest three frequencies (2, 5 and 15 seconds) and negative and significant for all but one of other frequencies, indicating that tick-time sampling is better than calendar-time sampling (has smaller average distance from the true QV) for all but the very highest frequencies. Further, Table 6 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.

[INSERT TABLE 6 ABOUT HERE]

4.2.3 Quote prices vs. 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 per day to the number of trades per day. I regress the difference in the accuracy of a quote-price RV and trade-price RV, with the same calendar-time sampling frequency, on a constant and the lagged ratio of the number of quotes to the number of trades. I do this for each of the frequencies considered in the earlier sections, and present the results in Table 7. The first column of Table 7 reveals that quote-price RV had larger average distance to the true QV than trade-price RV for all but two sampling frequencies, and for 10 out of 13 this difference is significant at the 0.05 level. However, the results of the test of conditional estimator accuracy reveal that quote-price RV improves relative to trade-price RV as the number of quote observations increases relative to the number of trades: 11 out of 13 slope coefficients are negative, and 10 of these are statistically significant. Results are very similar under the AR approximation, though with slightly reduced t-statistics. The ratio of quotes per day to trades per day for IBM has increased from around 0.5 in 1996 to around 2.5 in 2007, and may explain the sub-sample results in Table 5: as the relative number of quotes per day has increased, its relative accuracy has also increased. In the early part of the sample, quote-price RV was significantly less accurate than trade-price RV, however that difference vanishes in the last sub-sample, where quote and trade prices, of the same frequency, yield approximately equally accurate RV estimators[13].

[INSERT TABLE 7 ABOUT HERE]

5 Conclusion

This paper considers the problem of ranking competing realised volatility (RV) estimators, motivated by the growing literature on nonparametric estimation of price variability using high-frequency data, see Andersen, *et al.* (2006) and Barndorff-Nielsen and Shephard (2007) for recent surveys. I provide conditions under which the relative average accuracy of competing estimators for the latent target variable can be consistently estimated from available data, using "large T" asymptotics, and show that existing tests from the forecast evaluation literature, such as Diebold-Mariano (1995), West (1996), White (2000), Hansen, *et al.* (2005), 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-Wolf (2005) tests reject the squared daily return and the 5-minute calendar-time RV in favour of an RV estimator using data sampled at between 15 seconds and 5 minutes. 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.

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

This paper leaves open several extensions. The first is to consider a wider collection of classes of estimators of quadratic variation, in addition to considering different sampling frequencies and sampling schemes. Such a "horse race" may provide guidance on the type estimator that tends to work best for particular assets. A step in this direction is considered in Patton and Sheppard (2008), who look at combinations of realised volatility estimators. A second extension of the results in this paper is to comparisons of estimators of the entire covariance matrix. Such comparisons are perhaps more relevant than comparisons of individual variances and covariances, given that these components are usually used together as a covariance matrix (and thus must satisfy conditions to ensure that the matrix is positive semi-definite). For this application, the covariance matrix pseudo-distance measures proposed in Patton and Sheppard (2006) may prove useful, when combined with a random walk or a vector AR approximation for the latent integrated covariance matrix.

6 Appendix: Proofs

Additional assumptions used in parts of the proofs below:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

However it is *not* true that $Corr\left[C\left(X_{it}\right) - C\left(\theta_t\right), \tilde{\theta}_t - \theta_t | \mathcal{F}_{t-1}\right] = 0$ for all empirically relevant combinations of RV estimators and volatility proxies. In fact, if $X_{it} = \tilde{\theta}_t$ and $L = MSE$, a very natural case to consider, then $C(z) = -z$ and $Corr\left[C\left(X_{it}\right) - C\left(\theta_t\right), \tilde{\theta}_t - \theta_t | \mathcal{F}_{t-1}\right] = Corr\left[\theta_t - \tilde{\theta}_t, \tilde{\theta}_t - \theta_t | \mathcal{F}_{t-1}\right] = -1$. In general, we should expect $Corr\left[C\left(X_{it}\right) - C\left(\theta_t\right), \tilde{\theta}_t - \theta_t | \mathcal{F}_{t-1}\right] \neq 0$. This is the correlation between the error in $\tilde{\theta}_t$ and something similar to the "generalised forecast error", see Granger (1999) or Patton and Timmermann (2008), of X_{it}. If the proxy, $\tilde{\theta}_t$, and the RV estimators, X_{it}, use the same or similar data then their errors will generally be correlated and this zero correlation restriction will not hold, and thus $E\left[\left(C\left(X_{it}\right) - C\left(\theta_t\right)\right)\left(\tilde{\theta}_t - \theta_t\right)\right] \neq 0$, which breaks the asymptotic equivalence of the ranking obtained using $\tilde{\theta}_t$ with that using θ_t. ■

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

(b): Note that

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}\Delta L\left(Y_t,\mathbf{X}_t\right)-E\left[\Delta L\left(\theta_t,\mathbf{X}_t\right)\right] \\
= & \frac{1}{T}\sum_{t=1}^{T}\Delta L\left(\theta_t,\mathbf{X}_t\right)-E\left[\Delta L\left(\theta_t,\mathbf{X}_t\right)\right]+\frac{1}{T}\sum_{t=1}^{T}\left\{\Delta L\left(Y_t,\mathbf{X}_t\right)-\Delta L\left(\theta_t,\mathbf{X}_t\right)\right\} \\
= & \frac{1}{T}\sum_{t=1}^{T}\Delta L\left(\theta_t,\mathbf{X}_t\right)-E\left[\Delta L\left(\theta_t,\mathbf{X}_t\right)\right]+\frac{1}{T}\sum_{t=1}^{T}\Delta C\left(\mathbf{X}_t\right)\left(Y_t-\theta_t\right) \\
\equiv & \bar{\mathbf{A}}_T-E\left[\mathbf{A}_t\right] \\
\text{where } \mathbf{A}_t \equiv & \left[\Delta L\left(\theta_t,\mathbf{X}_t\right)',\Delta C\left(\mathbf{X}_t\right)'\left(Y_t-\theta_t\right)\right]'
\end{aligned}
$$

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

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

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

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

where $\Omega_1\equiv\boldsymbol{\iota}'V_A\boldsymbol{\iota}$. It should be noted that assumptions A1 and A2 can hold despite the random walk assumption (T1), if θ_t and X_{it} obey a some form of cointegration, linked to the distance measure employed. If MSE is employed, T1, A1 and A2 require that these variables obey standard linear cointegration, with cointegrating vector $[1,-1]$. For other distance measures a form of non-linear cointegration must hold.

(c): Follows directly from Theorem 3 of Politis and Romano (1994), under the additional assumption B1. ■

Proof of Proposition 3. (a): Using the second-order mean-value expansion of the loss function from the proof of Proposition 4, we obtain $E\left[\Delta L\left(Y_t,\mathbf{X}_t\right)\right]=E\left[\Delta L\left(\theta_t,\mathbf{X}_t\right)\right]+\boldsymbol{\beta}$, where $\boldsymbol{\beta}\equiv E\left[\Delta C\left(\mathbf{X}_t\right)\left(Y_t-\theta_t\right)\right]=\sum_{j=1}^{J}\omega_jE\left[\Delta C\left(\mathbf{X}_t\right)\left(\tilde{\theta}_{t+j}-\theta_t\right)\right]=\sum_{j=1}^{J}\omega_jE\left[\Delta C\left(\mathbf{X}_t\right)\left(\theta_{t+j}-\theta_t\right)\right]=\sum_{j=1}^{J}\omega_jE\left[\Delta C\left(\mathbf{X}_t\right)\left(E_t\left[\theta_{t+j}\right]-\theta_t\right)\right]$ under P1 and P2. Allowing for $J>1$ requires computing

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

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

$$\begin{aligned}
\text{redefine as} \quad P\mathbf{Z}_t &= \mathbf{Q}_0 + Q_1\mathbf{Z}_{t-1} + \mathbf{V}_t \\
\text{so} \quad \mathbf{Z}_t &= P^{-1}\mathbf{Q}_0 + P^{-1}Q_1\mathbf{Z}_{t-1} + P^{-1}\mathbf{V}_t \\
\text{with} \quad E\left[\mathbf{Z}_t\right] &= \left(I - P^{-1}Q_1\right)^{-1} P^{-1}\mathbf{Q}_0 \\
\text{and so } E_t\left[\mathbf{Z}_{t+j}\right] &= \left(I - P^{-1}Q_1\right)^{-1} P^{-1}\mathbf{Q}_0 + \left(P^{-1}Q_1\right)^j \left(\mathbf{Z}_t - \left(I - P^{-1}Q_1\right)^{-1} P^{-1}\mathbf{Q}_0\right) \\
&= \left\{\left(I - \left(P^{-1}Q_1\right)^j\right)\left(I - P^{-1}Q_1\right)^{-1} P^{-1}\mathbf{Q}_0\right\} + \left(P^{-1}Q_1\right)^j \mathbf{Z}_t \\
\text{and so} \quad E_t\left[\theta_{t+j}\right] &= g_0^{(j)} + \sum_{i=1}^{p} g_i^{(j)}\theta_{t+1-i}
\end{aligned}$$

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

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

which yields

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

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

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

Substituting in, we thus have

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

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

$$\begin{aligned}
\mathbf{B}_t &\equiv \Big[\Delta L\left(\theta_t,\mathbf{X}_t\right)',\Delta C\left(\mathbf{X}_t\right)',\Delta C\left(\mathbf{X}_t\right)'\tilde{\theta}_{t+1},...\Delta C\left(\mathbf{X}_t\right)'\tilde{\theta}_{t+J},\Delta C\left(\mathbf{X}_t\right)'\tilde{\theta}_{t-1},...\\
&\quad \Delta C\left(\mathbf{X}_t\right)'\tilde{\theta}_{t-p+1},\tilde{\theta}_t,\tilde{\theta}_t\tilde{\theta}_{t+1},...,\tilde{\theta}_t\tilde{\theta}_{t+2p}\Big]'.
\end{aligned} \tag{34}$$

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

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

$$\begin{aligned}
\Psi\boldsymbol{\phi} &= \boldsymbol{\psi}\\
\text{where }\ \Psi &\equiv \begin{bmatrix}\gamma_p & \gamma_{p-1} & \cdots & \gamma_1\\ \gamma_{p+1} & \gamma_p & \cdots & \gamma_2\\ \vdots & \vdots & \ddots & \vdots\\ \gamma_{2p-1} & \gamma_{2p-2} & \cdots & \gamma_p\end{bmatrix},\ \boldsymbol{\psi}=\left[\gamma_{p+1},\gamma_{p+2},...,\gamma_{2p}\right]',\ \boldsymbol{\phi}=\left[\phi_1,...,\phi_p\right]'
\end{aligned} \tag{35}$$

and by assumption T2 we can obtain $\hat{\boldsymbol{\phi}}=\hat{\Psi}^{-1}\hat{\boldsymbol{\psi}}$, where $\hat{\Psi}$ and $\hat{\boldsymbol{\psi}}$ are the equivalents of Ψ and $\boldsymbol{\psi}$ using sample autocovariances rather than population autocovariances. From $\hat{\boldsymbol{\phi}}$ we can obtain estimates of P, $\mathbf{Q}_0$, and Q_1 and thus estimates of the parameters $g_i^{(j)}$, for $i=0,1,...,p$ and $j=1,2,...,J$, from these we can compute the estimated bias term $\hat{\boldsymbol{\beta}}_T$. Given asymptotic normality of $\bar{\mathbf{B}}_T$ and the fact that $\frac{1}{T}\sum_{t=1}^{T}\Delta L\left(Y_t,\mathbf{X}_t\right)-\hat{\boldsymbol{\beta}}_T$ is a smooth function of the elements of $\bar{\mathbf{B}}_T$, we can then apply the delta method, see Lemma 2.5 of Hayashi (2000) for example, to obtain asymptotic normality of $\frac{1}{T}\sum_{t=1}^{T}\Delta L\left(Y_t,\mathbf{X}_t\right)-\hat{\boldsymbol{\beta}}_T$ and obtain its covariance matrix.

(c): Follows directly from Theorem 4 of Politis and Romano (1994), under the additional assumption B1. ■

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

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

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

$$\begin{aligned} \Delta L(Y_t, \mathbf{X}_t) &= \Delta L(\theta_t, \mathbf{X}_t) + \Delta C(\mathbf{X}_t)(Y_t - \theta_t) \\ \text{where } \Delta C(\mathbf{X}_t) &\equiv \begin{bmatrix} C(X_{1t}) - C(X_{2t}) \\ \vdots \\ C(X_{1t}) - C(X_{kt}) \end{bmatrix} \end{aligned}$$

Next, note:

$$\begin{aligned} E[\Delta C(\mathbf{X}_t)(Y_t - \theta_t)|\mathcal{G}_{t-1}] &= E\left[\Delta C(\mathbf{X}_t)\left(\sum_{i=1}^{J}\omega_i\tilde{\theta}_{t+i} - \theta_t\right)\middle|\mathcal{G}_{t-1}\right] \\ &= E\left[\Delta C(\mathbf{X}_t)\left(\sum_{i=1}^{J}\omega_i\sum_{j=1}^{i}\eta_{t+j} + \sum_{i=1}^{J}\omega_i\nu_{t+i}\right)\middle|\mathcal{G}_{t-1}\right] \\ &= E\left[\Delta C(\mathbf{X}_t)\left(\sum_{i=1}^{J}\omega_i\sum_{j=1}^{i}E\left[\eta_{t+j}|\mathcal{F}_t\right] + \sum_{i=1}^{J}\omega_i E\left[\nu_{t+i}|\mathcal{F}_t\right]\right)\middle|\mathcal{G}_{t-1}\right] = 0 \end{aligned}$$

by the law of iterated expectations, since $\mathcal{G}_{t-1} \subset \mathcal{F}_t$. This thus yields $E[\Delta L(\theta_t, \mathbf{X}_t)|\mathcal{G}_{t-1}] = E[\Delta L(Y_t, \mathbf{X}_t)|\mathcal{G}_{t-1}]$ as claimed.

(b): Using Exercise 5.21 of White (2001) for example, we have $\hat{D}_T^{-1/2}\sqrt{T}(\hat{\boldsymbol{\alpha}}_T - \tilde{\boldsymbol{\alpha}}) \to^d N(0, I)$, where $\hat{D}_T$ is given in the statement of the proposition.

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

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

$$
\begin{aligned}
E\left[\Delta L\left(Y_t, \mathbf{X}_t\right)\mathbf{Z}_{t-p}\right] - E\left[\Delta L\left(\theta_t, \mathbf{X}_t\right)\mathbf{Z}_{t-p}\right] &= \sum_{j=1}^{J}\omega_j E\left[\Delta C\left(\mathbf{X}_t\right)\left(\tilde{\theta}_{t+j}-\theta_t\right)\mathbf{Z}_{t-p}\right] \\
E\left[\Delta C\left(\mathbf{X}_t\right)\left(\tilde{\theta}_{t+j}-\theta_t\right)\mathbf{Z}_{t-p}\right] &= g_0^{(j)} E\left[\Delta C\left(\mathbf{X}_t\right)\mathbf{Z}_{t-p}\right] + \sum_{i=2}^{p} g_i^{(j)} E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1-i}\mathbf{Z}_{t-p}\right] \\
&\quad + \left(g_1^{(j)}-1\right) E\left[\Delta C\left(\mathbf{X}_t\right)\theta_t\mathbf{Z}_{t-p}\right] \\
\text{And } \; E\left[\Delta C\left(\mathbf{X}_t\right)\theta_t\mathbf{Z}_{t-p}\right] &= \frac{1}{\phi_1}E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1}\mathbf{Z}_{t-p}\right] - \frac{\phi_0}{\phi_1}E\left[\Delta C\left(\mathbf{X}_t\right)\mathbf{Z}_{t-p}\right] \\
&\quad - \sum_{i=2}^{p}\frac{\phi_i}{\phi_1}E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1-i}\mathbf{Z}_{t-p}\right]
\end{aligned}
$$

Pulling these results together we obtain:

$$
\begin{aligned}
& E\left[\Delta L\left(Y_t, \mathbf{X}_t\right)\mathbf{Z}_{t-p}\right] - E\left[\Delta L\left(\theta_t, \mathbf{X}_t\right)\mathbf{Z}_{t-p}\right] \\
= \;& \sum_{j=1}^{J}\omega_j E\left[\Delta C\left(\mathbf{X}_t\right)\left(\tilde{\theta}_{t+j}-\theta_t\right)\mathbf{Z}_{t-p}\right] \\
= \;& \sum_{j=1}^{J}\omega_j\{g_0^{(j)} E\left[\Delta C\left(\mathbf{X}_t\right)\mathbf{Z}_{t-p}\right] + \sum_{i=2}^{p} g_i^{(j)} E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1-i}\mathbf{Z}_{t-p}\right] \\
& + \left(g_1^{(j)}-1\right)\{\frac{1}{\phi_1}E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1}\mathbf{Z}_{t-p}\right] - \frac{\phi_0}{\phi_1}E\left[\Delta C\left(\mathbf{X}_t\right)\mathbf{Z}_{t-p}\right] - \sum_{i=2}^{p}\frac{\phi_i}{\phi_1}E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1-i}\mathbf{Z}_{t-p}\right]\}\} \\
= \;& E\left[\Delta C\left(\mathbf{X}_t\right)\mathbf{Z}_{t-p}\right]\left\{\sum_{j=1}^{J}\omega_j\left(g_0^{(j)} - g_1^{(j)}\frac{\phi_0}{\phi_1}\right) + \frac{\phi_0}{\phi_1}\right\} + \sum_{i=2}^{p} E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1-i}\mathbf{Z}_{t-p}\right] \times \\
& \left\{\sum_{j=1}^{J}\omega_j\left(g_i^{(j)} - g_1^{(j)}\frac{\phi_i}{\phi_1}\right) + \frac{\phi_i}{\phi_1}\right\} + E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1}\mathbf{Z}_{t-p}\right]\left\{\sum_{j=1}^{J}\omega_j\frac{g_1^{(j)}}{\phi_1} - \frac{1}{\phi_1}\right\} \\
\equiv \;& -\lambda_0 E\left[\Delta C\left(\mathbf{X}_t\right)\mathbf{Z}_{t-p}\right] - \sum_{i=2}^{p}\lambda_i E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1-i}\mathbf{Z}_{t-p}\right] - \lambda_1 E\left[\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1}\mathbf{Z}_{t-p}\right]
\end{aligned}
$$

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

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

$$
\begin{aligned}
\mathbf{D}_t \;\equiv\; & \left[\Delta L\left(\theta_t, \mathbf{X}_t\right)\mathbf{Z}_{t-p}', \Delta C\left(\mathbf{X}_t\right)\mathbf{Z}_{t-p}', \Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+1}\mathbf{Z}_{t-p}', ...\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t+J}\mathbf{Z}_{t-p}', ...\right. \\
& \left.\Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t-1}\mathbf{Z}_{t-p}', ..., \Delta C\left(\mathbf{X}_t\right)\tilde{\theta}_{t-p+1}\mathbf{Z}_{t-p}', \mathbf{Z}_t', \tilde{\theta}_t, \tilde{\theta}_t\tilde{\theta}_{t+1}, ..., \tilde{\theta}_t\tilde{\theta}_{t+2p}', \right]'
\end{aligned}
\tag{36}
$$

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

(c): Again follows directly from Theorem 4 of Politis and Romano (1994), under the additional assumption B1. ■

References

[1] Aït-Sahalia, Y., P. Mykland, and L. Zhang, 2005, How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise, *Review of Financial Studies*, 18, 351-416.

[2] Andersen, T.G., T. Bollerslev, 1998, Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts, *International Economic Review*, 39, 885-905.

[3] Andersen, T.G., T. Bollerslev, and N. Meddahi, 2004, Analytic Evaluation of Volatility Forecasts, *International Economic Review*, 45, 1079-1110.

[4] Andersen, T.G., T. Bollerslev, and N. Meddahi, 2007, Realized Volatility Forecasting and Market Microstructure Noise, working paper, Tanaka Business School, Imperial College London.

[5] Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys, 2000, Great Realizations, *Risk*, 13, 105-108.

[6] Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys, 2001a, The Distribution of Realized Exchange Rate Volatility, *Journal of the American Statistical Association*, 96, 42-55.

[7] Andersen, T.G., T. Bollerslev, F.X. Diebold, and H. Ebens, 2001b, The Distribution of Realized Stock Return Volatility, *Journal of Financial Economics*, 61, 43-76.

[8] Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys, 2003, Modeling and Forecasting Realized Volatility, *Econometrica*, 71, 579-626.

[9] Andersen, T.G., T. Bollerslev, P.F. Christoffersen, and F.X. Diebold, 2006, Volatility and Correlation Forecasting, in the *Handbook of Economic Forecasting*, G. Elliott, C.W.J. Granger and A. Timmermann eds., North Holland Press, Amsterdam.

[10] Bandi, F.M., and J.R. Russell, 2006a, Separating Microstructure Noise from Volatility, *Journal of Financial Economics*, 79, 655-692.

[11] Bandi, F.M., and J.R. Russell, 2006b, Comment on "Realized Variance and Microstructure Noise" by P.R. Hansen and A. Lunde, *Journal of Business and Economic Statistics*, 24, 167-173.

[12] Bandi, F.M., and J.R. Russell, 2006c, Market Microstructure Noise, Integrated Variance Estimators, and the Accuracy of Asymptotic Approximations, working paper, Graduate School of Business, University of Chicago.

[13] Bandi, F.M., and J.R. Russell, 2008, Microstructure Noise, Realized Variance, and Optimal Sampling, *Review of Economic Studies*, 75, 339-369.

[14] Bandi, F.M., J.R. Russell, and C. Yang 2007, Realized Volatility Forecasting in the Presence of Time-Varying Noise, working paper, Graduate School of Business, University of Chicago.

[15] Barndorff-Nielsen, O.E., and N. Shephard, 2002, Econometric Analysis of Realized Volatility and its use in Estimating Stochastic Volatility Models, *Journal of the Royal Statistical Society*, Series B, 64, 253-280.

[16] Barndorff-Nielsen, O.E., and N. Shephard, 2004, Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression and Correlation in Financial Economics, *Econometrica*, 72, 885-925.

[17] Barndorff-Nielsen, O.E., and N. Shephard, 2007, Variation, jumps, market frictions and high frequency data in financial econometrics, in *Advances in Economics and Econometrics. Theory and Applications*, Ninth World Congress, R. Blundell, P. Torsten and W.K. Newey eds., Econometric Society Monographs, Cambridge University Press, 328-372.

[18] Barndorff-Nielsen, O.E., B. Nielsen, N. Shephard, and C. Ysusi, 2004, Measuring and forecasting financial variability using realised variance, in A. Harvey, S.J. Koopman and N. Shephard (eds.), *State space and unobserved components models: Theory and Applications*, Cambridge University Press.

[19] Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde, and N. Shephard, 2008a, Designing Realised Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise, *Econometrica*, 76, 1481-1536.

[20] Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde, and N. Shephard, 2008b, Realized Kernels in Practice, *Econometrics Journal*, forthcoming.

[21] Bollerslev, T., R.F. Engle, and D.B. Nelson, 1994, ARCH Models, in the *Handbook of Econometrics*, R.F. Engle and D. McFadden eds, North Holland Press, Amsterdam.

[22] Christensen, K., and M. Podolskij, 2007, Realized range-based estimation of integrated variance, *Journal of Econometrics*, 141, 323-349.

[23] Diebold, F.X., and R.S. Mariano, 1995, Comparing Predictive Accuracy, *Journal of Business and Economic Statistics*, 13, 253-263.

[24] Easley, D. and M. O'Hara, 1992, Time and the Process of Security Price Adjustment, *Journal of Finance*, 47, 577-605.

[25] Engle, R.F., 2000, The Econometrics of Ultra High Frequency Data, *Econometrica*, 68, 1-22.

[26] Engle, R.F., and A.J. Patton, 2001, What Good is a Volatility Model?, *Quantitative Finance*, 1, 237-245.

[27] French, K.R., G.W. Schwert and R.F. Stambaugh, 1987, Expected Stock Returns and Volatility, *Journal of Financial Economics*, 19, 3-29.

[28] Gatheral, J., and R.C.A. Oomen, 2007, Zero-Intelligence Realized Variance Estimation, working paper, Department of Finance, Warwick Business School.

[29] Giacomini, R. and H. White, 2006, Tests of Conditional Predictive Ability, *Econometrica*, 74, 1545–1578.

[30] Gonçalves, S., and N. Meddahi, 2009, Bootstrapping Realized Volatility, *Econometrica*, 2009, 77, 283-306.

[31] Granger, C.W.J., 1999, Outline of Forecast Theory Using Generalized Cost Functions. *Spanish Economic Review*, 1, 161-173.

[32] Hansen, P.R., 2005, A Test for Superior Predictive Ability, *Journal of Business and Economic Statistics*, 23(4), 365-380.

[33] Hansen, P.R., and A. Lunde, 2005, A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?, *Journal of Applied Econometrics*, 20, 873-889.

[34] Hansen, P.R., and A. Lunde, 2006a, Realized Variance and Market Microstructure Noise, *Journal of Business and Economic Statistics*, 24, 127-161.

[35] Hansen, P.R., and A. Lunde, 2006b, Consistent Ranking of Volatility Models, *Journal of Econometrics*, 131, 97-121.

[36] Hansen, P.R., A. Lunde, and J.M. Nason, 2005, Model Confidence Sets for Forecasting Models, Federal Reserve Bank of Atlanta Working Paper 2005-7.

[37] Hayashi, F., 2000, *Econometrics*, Princeton University Press, New Jersey.

[38] Huang, X., and G. Tauchen, 2005, The Relative Contribution of Jumps to Total Price Variance, *Journal of Financial Econometrics*, 456-499.

[39] Kalnina, I., and O. Linton, 2007, Estimating Quadratic Variation Consistently in the presence of Correlated Measurement Error, *Journal of Econometrics*, forthcoming.

[40] Large, J., 2005, Estimating Quadratic Variation When Quoted Prices Change by a Constant Increment, *Journal of Econometrics*, forthcoming.

[41] Oomen, R.C.A., 2006, Properties of Realized Variance under Alternative Sampling Schemes, *Journal of Business and Economic Statistics*, 24, 219-237.

[42] Owens, J., and D. Steigerwald, 2007, Noise Reduced Realized Volatility: A Kalman Filter Approach, in D. Terrell, T. Fomby and R. C. Hill, eds., *Econometric Analysis of Financial and Economic Time Series*, Part A, Elsevier.

[43] Manganelli, S., 2005, Duration, Volume and Volatility Impact of Trades, *Journal of Financial Markets*, 8, 377-399.

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

[45] Merton, R.C., 1980, On Estimating the Expected Return on the Market: An Exploratory Investigation. *Journal of Financial Economics*, 8, 323-361.

[46] Newey, W.K., and K.D. West, 1987, A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica*, 55, 703-708.

[47] Patton, A.J., 2008, Volatility Forecast Comparison using Imperfect Volatility Proxies, *Journal of Econometrics*, forthcoming.

[48] Patton, A.J., and K. Sheppard, 2006, Evaluating Volatility Forecasts, in T.G. Andersen, R.A. Davis, J.-P. Kreiss and T. Mikosch (eds.) *Handbook of Financial Time Series*, Springer Verlag. Forthcoming.

[49] Patton, A.J., and K. Sheppard, 2008, Optimal Combinations of Realised Volatility Estimators, *International Journal of Forecasting*, forthcoming.

[50] Patton, A.J., and A. Timmermann, 2008, Generalized Forecast Errors, A Change of Measure, and Forecast Optimality Conditions, in *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, edited by T. Bollerslev, J.R. Russell and M.W. Watson, Oxford University Press, forthcoming.

[51] Politis, D.N., and J.P. Romano, 1994, The Stationary Bootstrap, *Journal of the American Statistical Association*, 89, 1303-1313.

[52] Poon, S.-H., and C.W.J. Granger, 2003, Forecasting Volatility in Financial Markets, *Journal of Economic Literature*, 41, 478-539.

[53] Romano, J.P., and M. Wolf, 2005, Stepwise multiple testing as formalized data snooping, *Econometrica*, 73, 1237-1282.

[54] West, K.D., 1996, Asymptotic Inference About Predictive Ability, *Econometrica*, 64, 1067-1084.

[55] White, H., 2000, A Reality Check for Data Snooping, *Econometrica*, 68, 1097-1126.

[56] White, H., 2001, *Asymptotic Theory for Econometricians*, Revised Edition, Academic Press, USA.

[57] Wright, J.H., 1999, Testing for a Unit Root in the Volatility of Asset Returns, *Journal of Applied Econometrics*, 14, 309-318.

[58] Zhang, L., 2006, Efficient Estimation of Stochastic Volatility using Noisy Observations: A Multi-Scale Approach, *Bernoulli*, 12, 1019-1043.

[59] Zhang, L., P.A. Mykland, and Y. Aït-Sahalia, 2005, A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data, *Journal of the American Statistical Association*, 100, 1394-1411.

[60] Zhou, B., 1996, High-Frequency Data and Volatility in Foreign-Exchange Rates, *Journal of Business and Economic Statistics*, 14, 45-52.

.

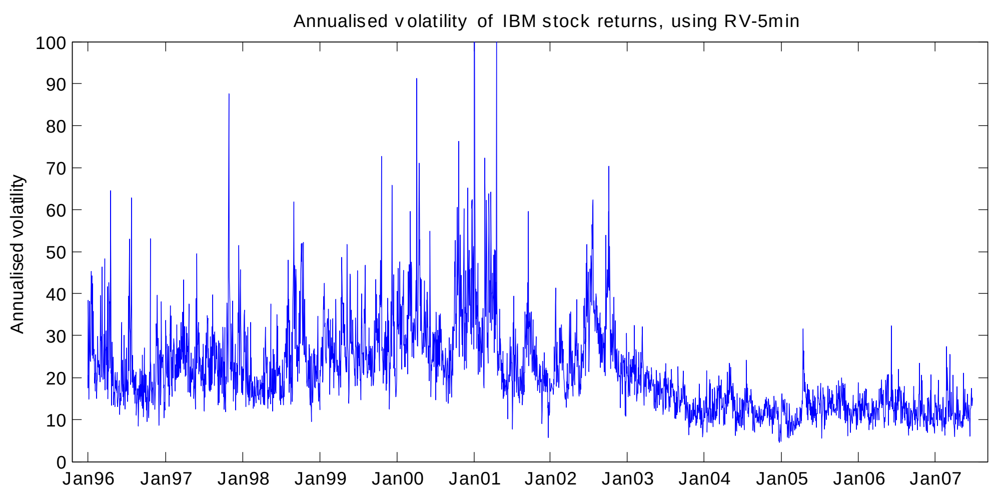

Figure 1: *IBM volatility over the period January 1996 to June 2007 (computed using realised volatility based on 5-minute calendar-time trade prices), annualised using the formula* $\sigma_t = \sqrt{252 \times RV_t}$.

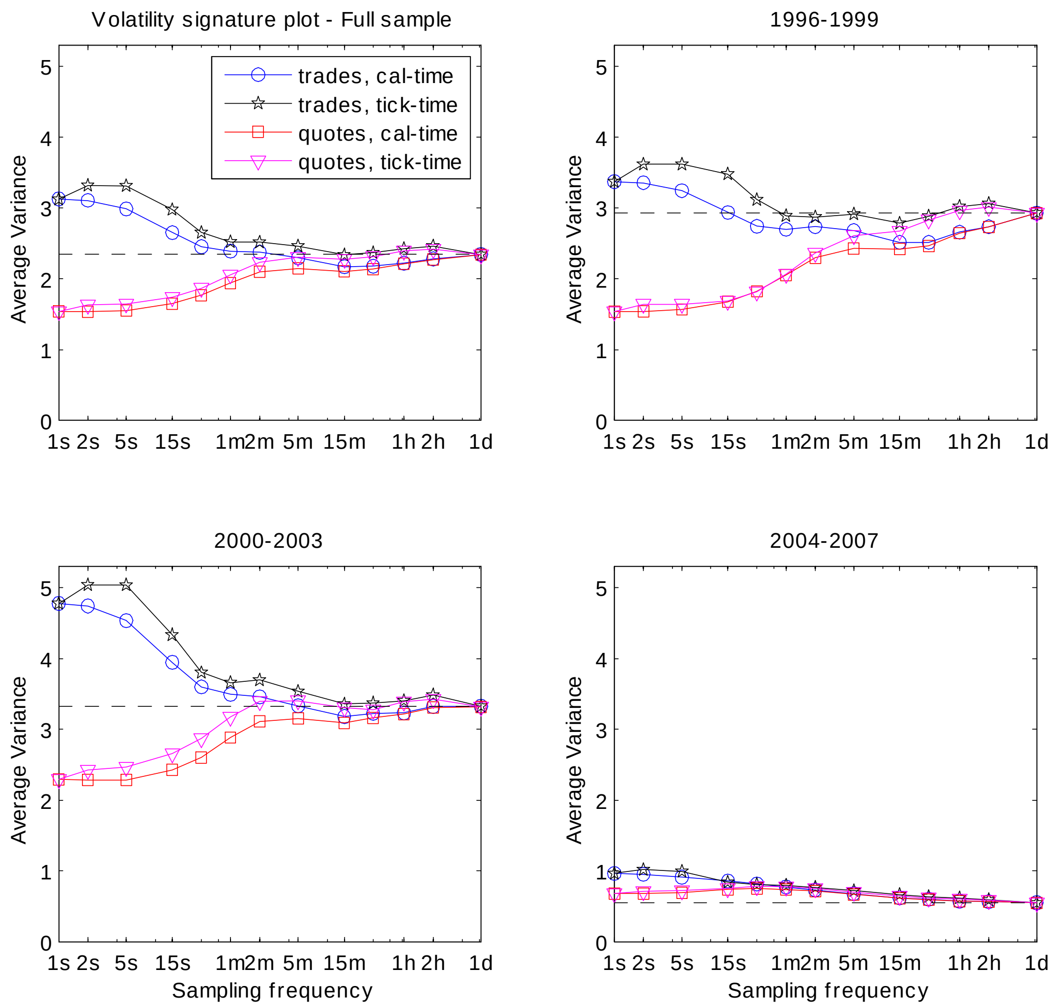

Figure 2: *'Volatility signature plots' for IBM, over the period January 1996 to June 2007, using 13 different sampling frequencies (from 1 second to 1 trade day), 2 different price series (trades and quotes) and 2 different sampling schemes (calendar-time and tick-time).*

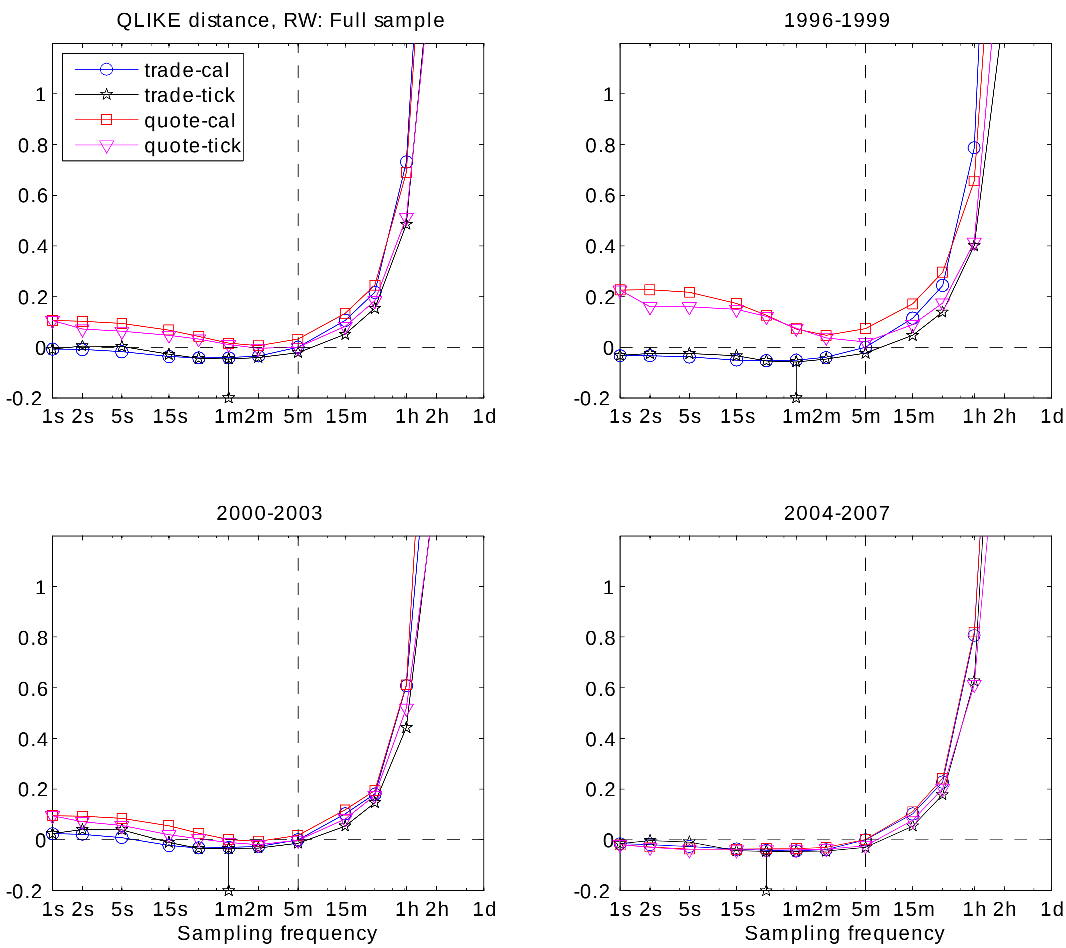

Figure 3: *Differences in average distance, estimated using a random walk approximation, for the 48 competing RV estimators, relative to 5-minute calendar-time RV on trade prices. A negative (positive) value indicates that the RV estimator is better (worse) than 5-minute calendar-time RV on trade prices. The estimator with the lowest average distance is marked with a vertical line down to the x-axis.*

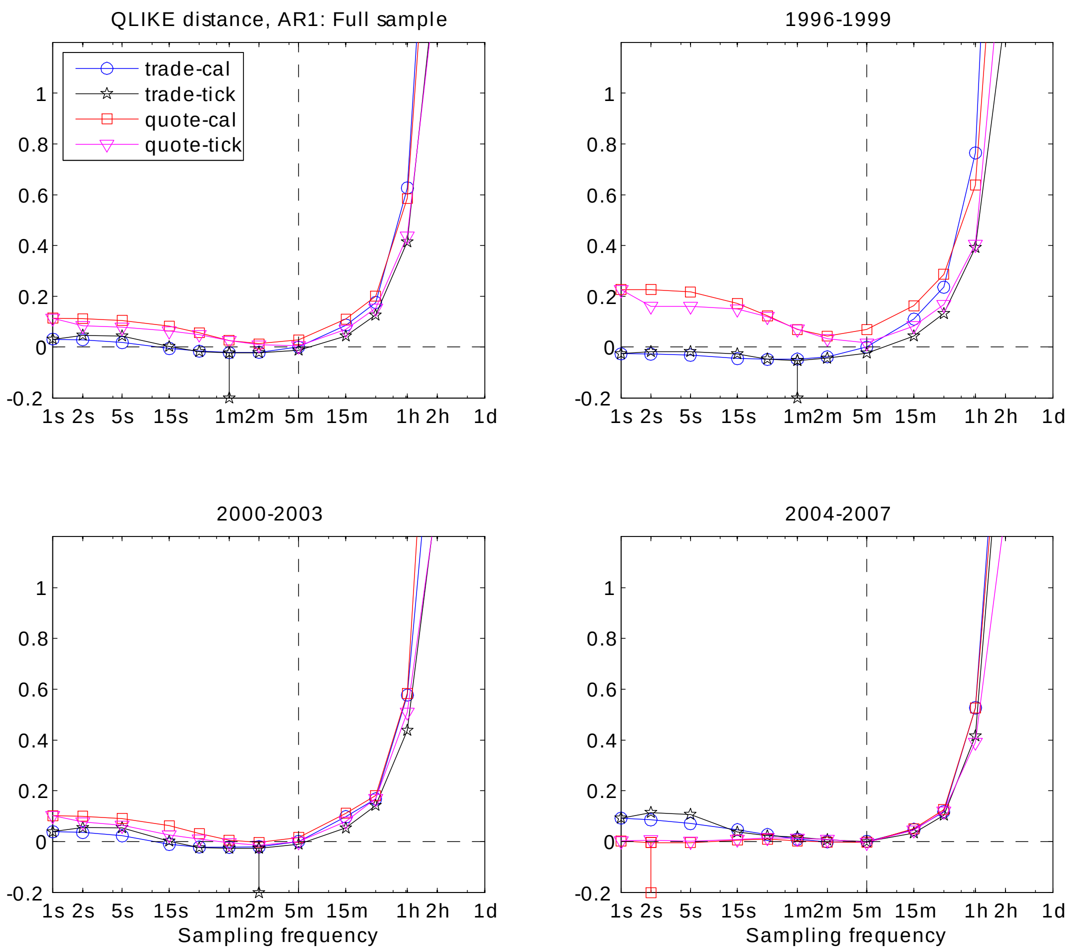

Figure 4: *Differences in average distance, estimated using an AR(1) approximation, for the 48 competing RV estimators, relative to 5-minute calendar-time RV on trade prices. A negative (positive) value indicates that the RV estimator is better (worse) than 5-minute calendar-time RV on trade prices. The estimator with the lowest average distance is marked with a vertical line down to the x-axis.*

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

			QV				*RV-30min*				*RV-daily*			
	QV^*		RW		AR		RW		AR		RW		AR	
T	*500*	*2500*	*500*	*2500*	*500*	*2500*	*500*	*2500*	*500*	*2500*	*500*	*2500*	*500*	*2500*
γ	**MSE**													
0.10	0.05	0.04	0.03	0.01	0.02	0.02	0.04	0.02	0.00	0.00	0.06	0.05	0.01	0.02
0.15	0.98	1.00	0.89	1.00	0.88	1.00	0.40	0.91	0.02	0.21	0.14	0.30	0.00	0.03
0.20	1.00	1.00	1.00	1.00	1.00	1.00	0.74	1.00	0.06	0.61	0.23	0.59	0.02	0.05
0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.56	0.82	0.60	1.00	0.06	0.10
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.70	0.84	0.89	1.00	0.07	0.14
	QLIKE													
0.10	0.02	0.03	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.02	0.01	0.01	0.01
0.15	0.22	0.29	0.12	0.19	0.10	0.18	0.08	0.18	0.06	0.19	0.04	0.06	0.02	0.01
0.20	0.38	0.61	0.26	0.53	0.23	0.55	0.21	0.48	0.13	0.43	0.10	0.23	0.02	0.02
0.50	0.66	0.96	0.60	0.94	0.60	0.95	0.56	0.90	0.38	0.77	0.29	0.70	0.07	0.08
1.00	0.77	0.99	0.76	0.98	0.76	0.98	0.74	0.97	0.47	0.82	0.54	0.87	0.10	0.12

Notes: This table presents the rejection frequencies for tests of equal accuracy of two competing RV estimators. 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)). All tests are conducted at the 0.05 level, based on 1000 draws from the stationary bootstrap. 1000 simulations are used for the $T = 500$ case, and 200 simulations are used for the $T = 2500$ case. The null hypothesis of equal average accuracy is satisfied in the first row of each panel, while in the other rows the second RV estimator has greater noise variance $(\gamma \equiv \sigma^2_{\zeta 2}/V\left[QV_t\right])$ than the first $(\sigma^2_{\zeta 1}/V\left[QV_t\right] = 0.10)$.

Table 2: Finite-sample size and power of conditional accuracy tests, RW approximation

	Unconditional				*Conditional - slope t-statistic*				*Conditional - joint test*			
		Volatility proxy				Volatility proxy				Volatility proxy		
λ	QV*	QV	RV-30min	RV-daily	QV*	QV	RV-30min	RV-daily	QV*	QV	RV-30min	RV-daily
						T=500						
0	0.03	0.03	0.03	0.03	0.02	0.02	0.03	0.02	0.01	0.01	0.02	0.02
0.1	0.09	0.10	0.10	0.05	0.04	0.04	0.04	0.02	0.11	0.11	0.10	0.04
0.2	0.29	0.28	0.29	0.16	0.06	0.06	0.06	0.02	0.40	0.39	0.36	0.13
0.4	0.72	0.72	0.70	0.56	0.08	0.10	0.10	0.04	0.80	0.80	0.78	0.55
0.8	0.98	0.98	0.97	0.93	0.57	0.55	0.46	0.18	0.95	0.98	0.97	0.93
						T=2500						
0	0.04	0.04	0.04	0.04	0.03	0.04	0.03	0.03	0.01	0.02	0.01	0.03
0.1	0.23	0.23	0.21	0.12	0.10	0.11	0.10	0.06	0.29	0.28	0.29	0.14
0.2	0.64	0.64	0.63	0.43	0.17	0.19	0.15	0.08	0.80	0.80	0.79	0.48
0.4	1.00	1.00	1.00	0.94	0.38	0.39	0.39	0.13	1.00	1.00	1.00	0.93
0.8	1.00	1.00	1.00	1.00	0.95	0.94	0.89	0.42	1.00	1.00	1.00	1.00

Notes: This table presents the rejection frequencies for tests of equal unconditional accuracy (Diebold-Mariano, 1995) and conditional accuracy (Giacomini-White, 2006) of two competing RV estimators, using the QLIKE pseudo-distance measure and the random walk approximation (assumption T1). The columns labelled QV* correspond these tests when the true QV is observable. The remaining columns present results based on Proposition 4, when the available volatility proxy has varying degrees of measurement error. All tests are conducted at the 0.05 level. 1000 simulations are used for the $T = 500$ case, and 200 simulations are used for the $T = 2500$ case. 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 at time t has time-varying bias equal to $-\lambda \times QV_{t-1}$.

Table 3: Finite-sample size and power of conditional accuracy tests, AR approximation

	Unconditional			*Conditional - slope t-statistic*			*Conditional - joint test*		
	Volatility proxy			Volatility proxy			Volatility proxy		
λ	QV	RV-30min	RV-daily	QV	RV-30min	RV-daily	QV	RV-30min	RV-daily
				T=500					
0	0.00	0.02	0.02	0.00	0.01	0.01	0.00	0.01	0.00
0.1	0.24	0.14	0.02	0.00	0.01	0.02	0.29	0.18	0.00
0.2	0.37	0.37	0.07	0.02	0.02	0.03	0.44	0.45	0.04
0.4	0.55	0.58	0.19	0.11	0.08	0.05	0.63	0.61	0.13
0.8	0.91	0.90	0.40	0.20	0.16	0.03	0.85	0.82	0.31
				T=2500					
0	0.01	0.02	0.01	0.00	0.01	0.00	0.00	0.00	0.00
0.1	0.11	0.10	0.03	0.02	0.02	0.00	0.10	0.12	0.01
0.2	0.51	0.49	0.08	0.06	0.07	0.01	0.58	0.59	0.06
0.4	0.97	0.95	0.45	0.20	0.24	0.00	0.95	0.93	0.41
0.8	1.00	0.99	0.88	0.92	0.86	0.22	0.98	0.99	0.80

Notes: This table presents the rejection frequencies for tests of equal unconditional accuracy (Diebold-Mariano, 1995) and conditional accuracy (Giacomini-White, 2006) of two competing RV estimators, using the QLIKE pseudo-distance measure and the AR(1) approximation (assumption T2). The columns labelled QV* correspond these tests when the true QV is observable. The remaining columns present results based on Proposition 5, when the available volatility proxy has varying degrees of measurement error. All tests are conducted at the 0.05 level. 1000 simulations are used for the $T = 500$ case, and 200 simulations are used for the $T = 2500$ case. 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 at time t has time-varying bias equal to $-\lambda \times QV_{t-1}$.

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

	RW approximation				**AR approximation**			
	Trades		Quotes		Trades		Quotes	
Sampling frequency	Calendar	Tick	Calendar	Tick	Calendar	Tick	Calendar	Tick
1 sec	✓✓✓✓	——	✓✓✓✓	——	✓✓✓✓	——	✓✓✓✓	——
2 sec	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓
5 sec	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓
15 sec	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓
30 sec	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓
1 min	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓
2 min	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓
5 min	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓
15 min	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓
30 min	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓
1 hr	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓
2 hr	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓
1 day	★	——	∼∼∼∼	——	★	——	∼∼∼∼	——

Notes: This table presents the results of Romano-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 $\times$, and estimators that are not significantly different are marked with $\sim$. 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.

Table 5: Tests of equal RV accuracy, with 5-minute RV as benchmark

	RW approximation				**AR approximation**			
	Trades		Quotes		Trades		Quotes	
Sampling frequency	Calendar	Tick	Calendar	Tick	Calendar	Tick	Calendar	Tick
1 sec	∼∼∼∼	——	××× ∼	——	∼∼ × ∼	——	∼∼∼∼	——
2 sec	∼∼∼∼	∼∼ × ∼	××× ∼	××× ∼	∼∼ × ∼	∼∼ × ∼	× ∼∼∼	∼∼∼∼
5 sec	∼∼∼∼	∼∼ × ∼	×××✓	×××✓	∼∼ × ∼	∼∼ × ∼	× ∼∼∼	× ∼∼∼
15 sec	✓ ∼∼ ✓	✓ ∼ ✓✓	×××✓	×× ∼ ✓	∼∼∼∼	∼∼∼∼	× ∼∼∼	× ∼∼∼
30 sec	✓✓ ∼ ✓	✓✓✓✓	×××✓	×× ∼ ✓	∼∼∼∼	∼∼∼∼	× ∼∼∼	× ∼∼∼
1 min	✓✓✓✓	✓✓✓✓	×× ∼ ✓	∼ × ∼ ✓	∼∼∼∼	∼∼∼∼	× ∼∼∼	× ∼∼∼
2 min	✓✓✓✓	✓✓✓✓	∼ × ∼ ✓	∼ × ∼ ✓	∼∼∼∼	∼∼∼∼	× ∼∼∼	∼∼∼∼
5 min	★	✓ ∼ ✓✓	××× ∼	∼∼∼∼	★	∼∼∼∼	× ∼∼∼	∼∼∼∼
15 min	××××	××××	××××	××××	∼∼∼∼	∼∼∼∼	∼∼∼∼	∼∼∼∼
30 min	××××	××××	××××	××××	∼∼∼∼	∼∼∼∼	∼∼∼∼	∼∼∼∼
1 hr	××××	××××	××××	××××	∼∼∼∼	∼∼∼∼	∼∼∼∼	∼∼∼∼
2 hr	××××	××××	××××	××××	∼∼∼∼	∼∼∼∼	∼∼∼∼	∼∼∼∼
1 day	××××	——	××××	——	∼∼∼∼	——	∼∼∼∼	——

Notes: This table presents the results of Romano-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-minute trade prices sampled in calendar time, marked with an ★. Estimators that are significantly better than the benchmark, at the 0.05 level, are marked with ✓, estimators that are significantly worse than the benchmark are marked with ×, and estimators that are not significantly different are marked with ∼ . The four characters in each element of the above table correspond to the results of the test for the full sample (1996-2007), first sub-sample (1996-1999), second sub-sample (2000-2003) and third sub-sample (2004-2007) respectively.

Table 6: Tests of equal unconditional and conditional RV accuracy: Tick-time vs. calendar-time sampling

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

Notes: This table presents the estimated difference in average distance of tick-time and calendar-time RV estimators, $L\left(Y_t, RV_t^{\text{tick}(h)}\right) - L\left(Y_t, RV_t^{\text{cal}(h)}\right)$, either unconditionally, or via a regression on a constant and one-period lag of the log variance of intra-day trade durations, which is a measure of the irregularity of the arrivals of trade observations. A negative slope coefficient indicates that higher volatility of durations leads to an improvement in the accuracy of the tick-time RV estimator relative to a calendar-time RV estimator using the same (average) frequency. Trade prices are used for all RV estimators. The fourth and eighth columns present the p-values from a chi-squared test that both coefficients are equal to zero. Two approximations for the dynamics of QV are considered: a random walk (RW) and a first-order autoregression (AR). Inference under the RW approximation is based on Newey-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 7: Tests of equal unconditional conditional RV accuracy:
Quote prices vs. trade prices

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

Notes: This table presents the estimated difference in average distance of quote-price and trade-price RV estimators, $L\left(Y_t, RV_t^{\text{quote}(h)}\right) - L\left(Y_t, RV_t^{\text{trade}(h)}\right)$, either unconditionally, or via a regression on a constant and one-period lag of the ratio of the number of quote observations per day to the number of trade observations per day. A negative slope coefficient indicates that an increase in the number of quote observations relative to 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-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.