Data-Based Ranking of Realised Volatility Estimators*

Andrew J. Patton
University of Oxford

9 June 2007

Preliminary. Comments welcome.

Abstract

I propose a formal, data-based method for ranking realised variance (RV) estimators. In contrast, most rankings of RV estimators currently in the literature are either graphical in nature, most notably the "volatility signature plot", or rely on asymptotic approximations of the mean-squared errors of the estimators, or on simulations. The proposed method relies on the existence of a volatility proxy that is unbiased for the variable of interest, and satisfies a certain "zero correlation" condition. The zero correlation condition has some similarities with instrumental variables estimation. The volatility proxy must be unbiased but it does not need to be very precise; a simple and widely-available proxy for volatility is the daily squared return. From a small empirical application to IBM volatility estimation, I find that the daily squared return is significantly out-performed by an RV estimator based on intra-daily data, while simple RV estimators based on 5-minute returns (either in calendar time or in tick time) were not significantly out-performed by any of 32 competing RV estimators.

Keywords: realised variance, volatility forecasting, instrumental variables.

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

*I thank Roel Oomen, Neil Shephard and Kevin Sheppard for helpful comments and suggestions, without implicating them for any errors, and I thank Runquan Chen for excellent research assistance. Contact address: Department of Economics, University of Oxford, Manor Road, Oxford OX1 3UQ, United Kingdom. Email: andrew.patton@economics.ox.ac.uk. A complete draft of this paper will be available from http://www.economics.ox.ac.uk/members/andrew.patton/research.html.

1 Introduction

Most rankings of realised volatility (RV) estimators in the existing econometrics literature are either graphical in nature (as opposed to formal statistical tests), notably the "volatility signature plot" of Zhou (1996) and Andersen, et al. (2000) for example, or rely on asymptotic approximations of the mean-squared errors of the estimators, or on simulations. In this paper I propose a formal data-based ranking method, which does not rely on continuous record asymptotic approximations or on simulating a "realistic" description of real data. The proposed ranking method relies on the existence of a volatility proxy that is unbiased for the latent target variable, and satisfies an uncorrelatedness condition, described in detail below. This proxy must be unbiased but it may not need to be very precise. A simple and widely-available proxy for volatility is the daily squared return.

The use of a consistent, data-based ranking of RV estimators has numerous advantages over rankings obtained via asymptotic theory or simulations. Compared with the continuous record asymptotic approximations, it allows one to examine the *finite-sample* performance of these estimators, which can differ widely from their asymptotic performance, as noted by Bandi and Russell (2005). Furthermore, much of the asymptotic theory for RV estimators in the presence of market microstructure noise relies on very specific assumptions about the noise process. For example, Hansen and Lunde (2006a) consider noise that is *iid* and additive to the efficient log-price process, or that is mean zero and covariance stationary. Zhang, *et al.* (2005) also consider the *iid* noise case, as do Barndorff-Nielsen, *et al.* (2006) and Bandi and Russell (2005). Finally, a data-based ranking method allows one to make comparisons that are difficult using existing theory: estimators based on trade prices versus quote prices; estimators defined using calendar-time sampling versus tick-time sampling (see Oomen, 2006, for a comparison of these estimators); or estimators based on very different assumptions, such as the multi-scale sub-sampled realised variance estimator of Zhang (2006) versus the 'alternation' estimator of Large (2005).

A data-based ranking has the obvious advantage over a simulation-based approach in that the latter requires a complete specification of the data generating process, and results obtained under one specification/parameterisation need not necessarily hold more generally. The data-based approach presented here allows one to answer the question of immediate interest to users of RV estimators: "which estimator works best on *my* asset return series, for *my* sample period?" I provide conditions sufficient for a consistent estimate of the average difference in distance between an RV estimator and the latent target variable to be obtained, which then allows one to use standard tests for forecast comparisons, such as those of Diebold and Mariano (1995), White (2000) and Hansen (2005). These tests rely on the usual large sample ($T \to \infty$) asymptotics, but do not rely on continuous record ($m \to \infty$, where m is the number of intra-period observations) asymptotics. This allows for formal comparisons of estimators that differ only in their sampling frequency.

1.1 Notation

θ_t is the latent target variable. I assume that θ_t is $\mathcal{F}_t$-measurable, though it is not observable to the econometrician. For the remainder of the paper I assume that θ_t is a scalar; I discuss the extension to vector target variables in the conclusion.

X_{it}, $i = 1, 2, ..., n$ are the realised volatility estimators to be ranked. Often these will be the same estimator applied to data sampled at different frequencies, for example 1-minute returns vs. 30-minute returns. They could also be estimators based on different measures of the price: trades vs. mid-quotes, for example, or different sampling schemes, such as calendar-time sampling vs. tick-time sampling.

In order to rank the competing estimators we need some measure of distance from the estimator, X_{it}, to the target variable, θ_t. In rankings of estimators based on asymptotic approximations this distance is usually the mean-squared error (MSE). When the two estimators are both consistent this reduces to comparing the asymptotic variances of the two estimators. Barndorff-Nielsen, *et al.* (2006) provide a detailed study of the asymptotic accuracy of a wide variety of 'kernel-based' realised volatility estimators, Hansen and Lunde (2006a) study the asymptotic MSE of a variety of estimators under different assumptions on the microstructure noise, while Bandi and Russell (2005) study the finite-sample MSE of some kernel-based realised volatility estimators under the assumption of *iid* microstructure noise. The extensive simulation study of Gatheral and Oomen (2007) also uses MSE to measure the distance between the estimator and the target variable.

I will consider ranking RV estimators using the average distance between the estimator and the target variable, using the general class of pseudo-distance measures proposed in Patton (2006):

$$E\left[L\left(\theta_t, X_{it}\right)\right] \quad \text{vs.} \quad E\left[L\left(\theta_t, X_{jt}\right)\right] \tag{1}$$

$$\text{where } L(\theta, X) = \tilde{C}(X) - \tilde{C}(\theta) + C(X)(\theta - X) \tag{2}$$

$$\text{and } \tilde{C}(z) \equiv \int C(z)\, dz$$

with C being some decreasing, twice-differentiable function. When both θ_t and X_{it} are strictly positive *a.s.*, the parametric family loss functions obtained from:

$$C(z; b) = \begin{cases} -(b+1)^{-1} z^{b+1}, & \text{for } b \notin \{-1, -2\} \\ -\log z, & \text{for } b = -1 \\ z^{-1}, & \text{for } b = -2 \end{cases}$$

$$\text{and so } \tilde{C}(z; b) = \begin{cases} -(b+1)^{-1}(b+2)^{-1} z^{b+2}, & \text{for } b \notin \{-1, -2\} \\ z - z\log z, & \text{for } b = -1 \\ \log z, & \text{for } b = -2 \end{cases}$$

has some attractive properties, see Patton (2006). This class nests MSE as a special case ($b = 0$, and so $C(z) = -z$) and the popular "QLIKE" loss function ($b = -2$), up to location and scale constants in both cases. More generally, the "shape" parameter b affects the penalty applied to over-estimation compared with under-estimation. This class is well-defined when both θ and X are almost surely strictly positive, which is a reasonable assumption in applications involving realised variance.

When either θ_t or X_{it} can be negative a different parametric family of pseudo-distance measures will be required. In all cases, the MSE distance measure can of course be employed.

Our interest is in measuring the average distance between the estimator and the latent target variable. I will obtain a consistent estimator of this quantity by employing a *proxy* or *instrument* for θ_t, denoted Y_t. The proxy must be observable by the econometrician, for the ranking to be "data-based", and must satisfy certain unbiasedness and zero correlation conditions. Deriving these conditions and finding a proxy that satisfies them is the main technical challenge in this paper.

The method I propose below relies on the presence of a potentially noisy but conditionally unbiased proxy for the latent target variable. For many assets the squared daily return can reasonably be assumed to be conditionally unbiased: the expected 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 RV 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 (2006) 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 even in the presence of jumps is available, however, the methods presented below apply directly.

2 Relation to the ranking of volatility forecasts

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 realised volatility estimators. Hansen and Lunde (2006b) and Patton (2006) show that if:

$$E\left[Y_t|\mathcal{F}_{t-1}\right] = \theta_t$$

(i.e., the proxy is conditionally unbiased for θ_t) then for any pseudo-distance measure in the class in equation (2) rankings based on the proxy are ($T-$asymptotically) equivalent to rankings based on the true unobservable target variable, assuming that the expectations exist. That is,

$$E\left[L\left(\theta_t, X_{1t}\right)\right] \lesseqqgtr E\left[L\left(\theta_t, X_{2t}\right)\right] \Leftrightarrow E\left[L\left(Y_t, X_{1t}\right)\right] \lesseqqgtr E\left[L\left(Y_t, X_{2t}\right)\right] \tag{3}$$

However, this result does not go through when (X_{1t}, X_{2t}) are RV estimators rather than a volatility forecasts. To see this, consider a mean-value expansion of the pseudo-distance measure $L\left(Y_t, X_t\right)$ given in equation (2) around (θ_t, X_t)

$$\begin{aligned}
L\left(Y_t, X_t\right) &= L\left(\theta_t, X_t\right) + \frac{\partial L\left(\theta_t, X_t\right)}{\partial \theta}\left(Y_t - \theta_t\right) + \frac{1}{2}\frac{\partial^2 L\left(\ddot{\theta}_t, X_t\right)}{\partial \theta^2}\left(Y_t - \theta_t\right)^2 \\
&= L\left(\theta_t, X_t\right) + \left(C\left(X_t\right) - C\left(\theta_t\right)\right)\left(Y_t - \theta_t\right) - \frac{1}{2}C'\left(\ddot{\theta}_t\right)\left(Y_t - \theta_t\right)^2 \\
\text{where } \ddot{\theta}_t &= \lambda_t \theta_t + \left(1 - \lambda_t\right) Y_t, \quad \lambda_t \in [0,1] \\
\text{then } E_{t-1}\left[L\left(Y_t, X_t\right)\right] &= E_{t-1}\left[L\left(\theta_t, X_t\right)\right] + E_{t-1}\left[\left(C\left(X_t\right) - C\left(\theta_t\right)\right)\left(Y_t - \theta_t\right)\right] \\
&\quad -\frac{1}{2}E_{t-1}\left[C'\left(\ddot{\theta}_t\right)\left(Y_t - \theta_t\right)^2\right]
\end{aligned} \tag{4}$$

The third term in equation (4) does not depend on X_t, and so will not affect the ranking of (X_{1t}, X_{2t}). For the ranking obtained using Y_t to be the same as that obtained using θ_t we need to show that the second term equals zero:

$$E_{t-1}\left[\left(C\left(X_t\right) - C\left(\theta_t\right)\right) \cdot \left(Y_t - \theta_t\right)\right] = 0$$

In the standard case, X_t is a volatility *forecast* and θ_t is the conditional variance, and so is $\mathcal{F}_{t-1}$-measurable, which allows:

$$E_{t-1}\left[\left(C\left(X_t\right) - C\left(\theta_t\right)\right) \cdot \left(Y_t - \theta_t\right)\right] = \left(C\left(X_t\right) - C\left(\theta_t\right)\right) \cdot E_{t-1}\left[Y_t - \theta_t\right] = 0$$

by the conditional unbiasedness of Y_t. However, when X_t is a realised volatility estimator and θ_t is the integrated variance or quadratic variation we have $(X_t, \theta_t) \in \mathcal{F}_t$ but $(X_t, \theta_t) \notin \mathcal{F}_{t-1}$, and so we cannot take the first term above out of the expectation. In short, the fact that the realised variance estimator of the target variable for day t is only available at the *end* of day t rules out the direct application of established results for volatility forecast comparison.

If we could assume that

$$Corr_{t-1}\left[C\left(X_{it}\right) - C\left(\theta_t\right), Y_t - \theta_t\right] = 0 \;\; \forall \; i$$

in addition to $E\left[Y_t | \mathcal{F}_{t-1}, \theta_t\right] = \theta_t$, then we would have

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

But it is not likely that $Corr_{t-1}\left[C\left(X_{it}\right) - C\left(\theta_t\right), Y_t - \theta_t\right] = 0$ for all combinations of RV estimators and volatility proxies. For example, if $X_{it} = Y_t$ and $L = MSE$, then

$$\begin{aligned} C\left(z\right) &= -z \\ \text{so } \; Corr_{t-1}\left[C\left(X_{it}\right) - C\left(\theta_t\right), Y_t - \theta_t\right] &= Corr_{t-1}\left[\theta_t - Y_t, Y_t - \theta_t\right] = -1 \end{aligned}$$

Thus this correlation will in fact equal -1! More generally, we would expect this correlation to be non-zero. It is the correlation between the error in Y_t and something similar to the "generalised forecast error", see Granger (1999) or Patton and Timmermann (2003), of X_t. If the proxy, Y_t, and the RV estimator, X_t, are both using the same or similar information sets then their errors will generally be correlated and this zero correlation restriction will not hold. This reveals the similarity of this problem to instrumental variables estimation: ignoring the correlation between the error in the RV estimator and the error in the proxy leads to invalid inference.

3 Data-based ranking of RV estimators

I present results under two broad sets of assumptions: the first allows for general behaviour in the target variable, θ_t, but restricts the behaviour of the RV estimators, X_{it}. The second set of assumptions allows for general behavior of the RV estimators, at the cost of imposing some restrictions on the behaviour of the target variable. We present both sets of results as in different applications one set of assumptions may be more palatable than the other.

3.1 Rankings based on assumptions about the RV estimator bias

This section presents results for data-based ranking of RV estimators that hold when we can assume that the time series behaviour of the bias in the RV estimators satisfies restriction given below.

Assumption T1: θ_t is a mean stationary process.

Assumption P1: Y_t is a mean stationary process with $E[Y_t] = E[\theta_t]$.

Assumption P2: $Y_t \in \mathcal{F}_{t-1}$.

Assumption R1: $E[X_{it}|\theta_t, \mathcal{F}_{t-1}] = \theta_t + c_i\theta_t^k \ \forall\ i$, where k is known, and $\max_i \ |c_i| < \infty$.

The first two of these assumptions are standard, with only unconditional unbiasedness of Y_t required (rather than conditional unbiasedness). Assumption P2 requires that the proxy is measurable at time $t-1$, which is non-standard. We would usually consider a proxy for θ_t as being something measured on day t, such as the squared returns from day t. Assumption P2 suggests instead to use the first lag of the daily squared return, or longer lags, or perhaps combinations of lags. (We will consider optimal choices of proxies below.) The result below shows that using lagged squared returns can be useful in obtaining a data-based ranking of RV estimators.

Assumption R1 is the key assumption for this result. It requires that the bias in the RV estimators be proportional to some power of the target variable, with a common power but potentially different proportionality constants. This nests the interesting special cases where all RV estimators are unbiased ($c_i = 0\ \forall\ i$). Also nested is the case where all RV estimators have some biases that are constant through time ($k = 0$) but which can differ across estimators ($c_i \neq c_j$). This is relevant, for example, if the observed price is equal to the true efficient price plus some covariance stationary or *iid* noise, see Hansen and Lunde (2006a, Theorem 1 and Lemma 2). Finally, assumption R1 also allows the biases in the RV estimators to be proportional to some power of θ_t, for example θ_t^2. This might be of interest as in many cases asymptotic variance of many RV estimators is related to integrated quarticity, which is in turn related to the square of integrated variance, see Barndorff-Nielsen and Shephard (2004) for example. Alternatively, if the noise in the observed price is proportional to the observed price, and the RV estimator is based on tick-time sampling, then the bias is proportional to θ_t, see Hansen and Lunde (2006a, Example 1).

Proposition 1 *Let assumptions T1, P1, P2 and R1 hold. Then if* $(\theta_t, Y_t, X_{1t}, X_{2t})'$ *is strictly positive a.s.,*

$$\begin{aligned} E\left[L\left(\theta_t, X_{1t}; b\right)\right] - E\left[L\left(\theta_t, X_{2t}; b\right)\right] &= E\left[L\left(Y_t, X_{1t}; b\right)\right] - E\left[L\left(Y_t, X_{2t}; b\right)\right], \quad \textit{if } b = k = 0 \\ &\approx E\left[L\left(Y_t, X_{1t}; b\right)\right] - E\left[L\left(Y_t, X_{2t}; b\right)\right], \quad \textit{if } b = -k \neq 0 \end{aligned}$$

for any two RV estimators, X_{1t} *and* X_{2t}*, where* k *is from assumption R1. If* $(\theta_t, Y_t, X_{1t}, X_{2t})'$ *can be negative this result holds for* $k = b = 0$.

All proofs are presented in the Appendix.

This proposition shows that if we can make some assumption about the time series behaviour of the bias in the competing RV estimators, then there exists a unique pseudo-distance measure from equation (2) that yields a feasible data-based ranking of RV estimators. For example, if the biases in the RV estimators are constant through time, then we can rank the RV estimators using MSE ($b = 0$). If the biases in the RV estimators are proportional to θ_t^2, then the QLIKE pseudo-distance measure ($b = -2$) should be used to rank the RV estimators. Note, importantly, that we do not need to assume anything about the behaviour of the biases as m (the number of intra-daily observations) varies. Different choices of m correspond to different RV estimators in this framework and no relation between the competing estimators is imposed.

3.2 Rankings based on a random walk assumption

The result from the previous section relied on a rather specific assumption about the time series properties of the biases in the competing RV estimators. In this section I do away with such assumptions by imposing some structure on the time series dynamics of the target variable, θ_t.

Numerous papers on the conditional variance (see Bollerslev, *et al.*, 1994, Engle and Patton, 2001, and Andersen, *et al.*, 2005 for example), or integrated variance (see Andersen, *et al.*, 2004 and 2005) have reported that these quantities are very persistent, close to being random walks. The popular RiskMetrics model, for example, is based on a unit root assumption for the conditional variance. Wright (1999) provides thorough evidence *against* the presence of a unit root in daily conditional variance for many stocks, however, despite this, it has proven to be a good approximation in many cases. Given this, consider the following assumption:

Assumption T2: $\theta_t = \theta_{t-1} + \eta_t$, with $E[\eta_t | \mathcal{F}_{t-1}] = 0$.

In the proof of the following proposition I need to strengthen the unconditional unbiasedness assumption in P1 to the standard conditional unbiasedness assumption. Let us denote the conditionally unbiased proxy as $\tilde{\theta}_t$, rather than Y_t, as below I will consider taking linear combinations of unbiased proxies to improve the power of tests in finite samples.

Assumption P1': $\tilde{\theta}_t = \theta_t + \nu_t$, with $E[\nu_t | \mathcal{F}_{t-1}, \theta_t] = 0$.

For the proposition below I again consider using a proxy for θ_t that is not measured on day t, but instead of considering lags of $\tilde{\theta}_t$ it turns out to be best to consider *leads* of $\tilde{\theta}_t$. The reason for this becomes clear in the proof.

Assumption P2': $Y_t = \sum_{i=1}^{J} \omega_i \tilde{\theta}_{t+i}$, where $1 \leq J < \infty$, $\omega_i \geq 0 \; \forall \, i$ and $\sum_{i=1}^{J} \omega_i = 1$.

Proposition 2 *Let assumptions T2, P1' and P2' hold. If* $(\theta_t, Y_t, X_{1t}, X_{2t})'$ *is strictly positive a.s., then*

$$E[L(\theta_t, X_{1t}; b)] - E[L(\theta_t, X_{2t}; b)] = E[L(Y_t, X_{1t}; b)] - E[L(Y_t, X_{2t}; b)]$$

for any two RV estimators, X_{1t} *and* X_{2t}*, and any* b*. If* $(\theta_t, Y_t, X_{1t}, X_{2t})'$ *can be negative this result holds for MSE loss (*$b = 0$*).*

In some ways the above result is substantially more general than that in Proposition 1. The assumption that θ_t follows a random walk allows us to leave the bias, if any, in the RV estimators completely unspecified: it can be constant, time-varying, a function of different powers of θ_t, or a function of other variables altogether. Furthermore, Proposition 2 shows that *any* pseudo-distance measure from the class in equation (2) may be used, according to the preferences of the user of the RV estimator. Note that for a formal RV comparison test to be implemented we will need certain moment conditions to be satisfied and this may restrict the choice of pseudo-distance measure.

An alternative motivation for the empirical approach suggested by Proposition 2 is based on the asymptotics of rolling window volatility estimators given in Foster and Nelson (1996), and used in Fleming, *et al.* (2001), amongst many other applications. Foster and Nelson show that, under some conditions, estimators such as those covered in assumption P2' converge to true (spot) variance as the length of the period $H = 1/m$ (one day, in our application in Section 5) goes to zero and as the number of intra-period observations, m, goes to infinity.

Before moving on, it is worth considering how the above proposition changes when the target variable is only "close to" a random walk. To that end, consider the following modification of the random walk assumption:

Assumption T2': $\theta_t = \mu + \phi(\theta_{t-1} - \mu) + \eta_t$, with $E[\eta_t|\mathcal{F}_{t-1}] = 0$, and $\mu \equiv \bar{\theta}\delta$, $\phi \equiv 1 - \delta$, where δ is a small positive constant.

Under this weaker assumption on the time series dynamics of the target variable I obtain the following result. For simplicity I restrict the proxy to be a simple lead of $\tilde{\theta}_t$.

Proposition 3 *Let assumptions T2' and P1' hold, and set $Y_t = \tilde{\theta}_{t+1}$. (i) Then*

$$
\begin{aligned}
E[L(\theta_t, X_{1t})] - E[L(\theta_t, X_{2t})] &= E[L(Y_t, X_{1t})] - E[L(Y_t, X_{2t})] \\
&\quad -\delta E[\theta_t (C(X_{1t}) - C(X_{2t}))] \\
&\quad +\bar{\theta}\delta^2 E[C(X_{1t}) - C(X_{2t})]
\end{aligned}
$$

for any two RV estimators, X_{1t} and X_{2t}.

(ii) If we further assume R1 and let $\epsilon \equiv c_2 - c_1$, then if $(\theta_t, Y_t, X_{1t}, X_{2t})'$ is strictly positive a.s. we have

$$
\begin{aligned}
\delta E[\theta_t (C(X_{1t}; b) - C(X_{2t}; b))] &= \delta\epsilon E\left[\theta_t^{k+1}\right], \quad \text{for } b = 0 \\
&\approx \delta\epsilon E\left[\theta_t^{k+1+b}\right], \quad \text{for } b \neq 0
\end{aligned}
$$

$$
\begin{aligned}
\text{and} \quad \bar{\theta}\delta^2 E[C(X_t; b) - C(\theta_t; b)] &= \bar{\theta}\delta^2\epsilon E\left[\theta_t^{k}\right], \quad \text{for } b = 0 \\
&\approx \bar{\theta}\delta^2\epsilon E\left[\theta_t^{k+b}\right], \quad \text{for } b \neq 0
\end{aligned}
$$

When $(\theta_t, Y_t, X_{1t}, X_{2t})'$ can be negative the result for $b = 0$ holds.

The first part of Proposition 3 shows explicitly the extra terms that appear when the target variable follows an AR(1) rather than a random walk. The second part of the proposition provides some idea of the magnitudes of these terms as a function of δ, which measures how close to a random walk the target variable is, and ϵ, which measures the difference in the proportionality constants in the biases of the two RV estimators. Empirically it is widely found that δ is positive but small. Thus the third term in part (i) is $O\left(\delta^2\right)$ and may be negligible. The second term is $O(\delta)$, and becomes $O(\delta\epsilon)$ when we impose some structure on the biases in the RV estimators. If δ is small, and we think ϵ is small, then this term will also be negligible.

Proposition 3 provides some reassurance of the empirical usefulness of the ranking method suggested by Proposition 2: if the target variable is close to a random walk, and/or the RV estimators being compared have similar biases, then ranking RV estimators by using a lead of a conditionally unbiased proxy for θ_t in conjunction with a pseudo-distance measure from equation (2) will yield the same ranking as if θ_t was directly observable.

Proposition 2 above suggests the use of a convex combination of leads of $\tilde{\theta}_t$, but gives no guidance on how many leads, J, to consider or on the appropriate weights to apply to each lead individually. While the weighting function could theoretically have $J-1$ free parameters (it must sum to one, pinning down the J^{th} weight) let us simplify the problem and consider only equally-weighted proxies. In this case, the problem reduces to choosing J, the number of leads to combine.

Proposition 4 *Let P1', P2' and T2' hold. Then, imposing $\omega_i = J^{-1}\ \forall\ i = 1,2,...,J$, the variance of the error in the proxy for a given value of J is*

$$V\left[Y_t - \theta_t\right] = \frac{1}{J}\sigma_\nu^2 + \frac{(J+1)(2J+1)}{6J}\sigma_\eta^2 + \left(1+\frac{1}{J}\right)\sigma_{\eta\nu}$$

The number of leads that minimises the variance of the measurement error in Y_t is given by

$$J^* = \sqrt{\frac{1+6k+6\rho\sqrt{k}}{2}}$$

$$\text{where } \psi \equiv \sigma_\nu^2/\sigma_\eta^2 \text{ and } \rho \equiv Corr\left[\nu_t, \eta_t\right]$$

When we constrain J^* to be an integer between 1 and 10000, the optimal values are:

J^*			ρ		
ψ	-0.9	-0.5	0	0.5	0.9
0.0001	1	1	1	1	1
0.1	1	1	1	1	1
1	1	1	2	2	3
10	5	5	6	6	6
100	17	17	17	18	18
10,000	172	173	173	174	174

These results reveal that the optimal integer values of J do not vary greatly with ρ, though they do change with ψ. When $\psi < 1$, it is intuitively clear that we should use only one lead of $\tilde{\theta}_t$, as in that case $\tilde{\theta}_t$ is a relatively accurate estimator of θ_t and the gains from smoothing are low. When $\psi \geq 1$, there is potentially some benefit to smoothing the proxy across a range of leads of $\tilde{\theta}_t$. Only for very large values of ψ do we average across more than a few leads of $\tilde{\theta}_t$.

It should be noted that the above result for the optimal value for J is very sensitive to the random walk assumption for θ_t: if θ_t is actually slowly mean-reverting then using leads of 100 or more periods will yield misleading results. In practice, it may be best to limit the value of J to be no more than 5 or 10 for daily data, depending on the estimated persistence in the latent target variable.

3.3 Rankings based on an AR(p) assumption

I now present the most generally applicable result of this paper, which allows the latent target variable to follow any stationary AR(p) process, subject to the first-order AR coefficient being different from zero. 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 such this generalisation of the result for random walks 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 for the fact that allowing for an ARMA model would theoretically involve estimating an infinite number, asymptotically, of autocovariance-type quantities, which is likely to have unsatisfactory finite sample properties. In the next section I show via simulations that an AR(p) approximation to the process for daily integrated variance is barely distinguishable from the ARMA approximation, for one common stochastic volatility model.

This result requires a consistent estimator of the parameters of the AR(p) model for the latent target variable: we present such an estimator in the following lemma. This result follows from Baillie and Chung (2001), although I focus on an AR(p) model and autocovariances, rather than ARMA models and autocorrelations, which simplifies the estimator and allows for a closed-form expression.

Assumption T2": $\theta_t = \phi_0 + \sum_{i=1}^{p} \phi_i \theta_{t-i} + \eta_t$, with $E[\eta_t|\mathcal{F}_{t-1}] = 0$, $\phi_1 \neq 0$ and $[\phi_1, ..., \phi_p]'$ such that θ_t is covariance stationary.

Lemma 1 *Let assumptions P1' and T2" hold and define*

$$\begin{aligned}
\Phi &\equiv [\mu, \phi_1, \phi_2, ..., \phi_p]' \\
\hat{A}_T^{(k)} &\equiv \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & \hat{\gamma}_p & \hat{\gamma}_{p-1} & \cdots & \hat{\gamma}_1 \\ 0 & \hat{\gamma}_{p+1} & \hat{\gamma}_p & \cdots & \hat{\gamma}_2 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & \hat{\gamma}_{2p+k-1} & \hat{\gamma}_{2p+k-2} & \cdots & \hat{\gamma}_{p+k} \end{bmatrix} \\
\text{and } \; B_T^{(k)} &\equiv \begin{bmatrix} \hat{\mu}_T & \hat{\gamma}_{p+1} & \hat{\gamma}_{p+2} & \cdots & \hat{\gamma}_{2p+k} \end{bmatrix}', \; k \geq 0 \\
\text{where } \hat{\gamma}_j &\equiv \frac{1}{T-j} \sum_{t=j+1}^{T} \left(\tilde{\theta}_t - \hat{\mu}\right)\left(\tilde{\theta}_{t-j} - \hat{\mu}\right) \\
\hat{\mu}_T &\equiv \frac{1}{T} \sum_{t=1}^{T} \tilde{\theta}_t
\end{aligned}$$

and let

$$\hat{\Phi}_T = \left(\hat{A}_T^{(k)\prime} W_T \hat{A}_T^{(k)}\right)^{-1} \hat{A}_T^{(k)\prime} W_T \hat{B}_T^{(k)} \tag{5}$$

where $W_T \rightarrow^p W$*, a positive definite matrix of constants. Then*

$$\sqrt{T}\left(\hat{\Phi}_T - \Phi\right) \rightarrow^D \; N(0, V) \quad \text{as } T \rightarrow \infty.$$

When the order of the autoregression is greater than one, I also require the following assumption:

Assumption R2: X_{it} is independent of ν_{t-j} for all $j > 0$.

This assumption is mild given that almost all RV estimators (Barndorff-Nielsen, *et al.*, 2005, and Owen and Steigerwald, 2007, being two exceptions) are based solely on a single day of intra-day returns.

Proposition 5 *Let assumptions T2" and P1' hold, and let R2 hold if* $p > 1$. *(i) Then*

$$\begin{aligned}
E\left[L(\theta_t, X_{1t}; b) - L(\theta_t, X_{2t})\right] &= E\left[L\left(\tilde{\theta}_{t+1}, X_{1t}\right) - L\left(\tilde{\theta}_{t+1}, X_{2t}\right)\right] \\
&\quad + \frac{1-\phi_1}{\phi_1} E\left[\left(C(X_{1t}) - C(X_{2t})\right)\tilde{\theta}_{t+1}\right] \\
&\quad - \frac{\phi_0}{\phi_1} E\left[C(X_{1t}) - C(X_{2t})\right] \\
&\quad - \sum_{j=2}^{p} \frac{\phi_j}{\phi_1} E\left[\left(C(X_{1t}) - C(X_{2t})\right)\tilde{\theta}_{t+1-j}\right]
\end{aligned}$$

for any two RV estimators, X_{1t} and X_{2t}. If $(\theta_t, Y_t, X_{1t}, X_{2t})'$ is strictly positive a.s. this result holds for any b in the pseudo-distance measures presented in equation (2). If $(\theta_t, Y_t, X_{1t}, X_{2t})'$ may be negative this result holds for the squared distance measure, MSE $(b = 0)$.

(ii) Under standard regularity conditions, we have

$$\begin{aligned}
&\frac{1}{T-1}\sum_{t=1}^{T-1}\left\{L\left(\tilde{\theta}_{t+1}, X_{1t}; b\right) - L\left(\tilde{\theta}_{t+1}, X_{2t}\right)\right\} \\
&+\frac{1-\hat{\phi}_{1,T}}{\hat{\phi}_{1,T}}\cdot\frac{1}{T-1}\sum_{t=1}^{T-1}\left\{\left(C\left(X_{1t}\right) - C\left(X_{2t}\right)\right)\tilde{\theta}_{t+1}\right\} \\
&-\frac{\hat{\phi}_{0,T}}{\hat{\phi}_{1,T}}\frac{1}{T}\sum_{t=1}^{T}\left\{C\left(X_{1t}\right) - C\left(X_{2t}\right)\right\} \\
&-\sum_{j=2}^{p}\frac{\hat{\phi}_{j,T}}{\hat{\phi}_{1,T}}\frac{1}{T+1-j}\sum_{t=j}^{T}\left\{\left(C\left(X_{1t}\right) - C\left(X_{2t}\right)\right)\tilde{\theta}_{t+1-j}\right\} \\
\to^{p}\ & E\left[L\left(\theta_t, X_{1t}; b\right) - L\left(\theta_t, X_{2t}\right)\right], \quad \text{as } T \to \infty
\end{aligned}$$

where $\hat{\phi}_{i,T}$, $i = 0, 1, 2, ..., p$ are estimated using Lemma 1.

Proposition 5 relaxes the assumption of a random walk, at the cost of introducing a bias term to the expected loss computed using the proxy, even when the error in the proxy is uncorrelated with the error in the RV estimators (e.g., when a lead of $\tilde{\theta}_t$ is used). This bias term, however, can be consistently estimated under the assumption that the target variable follows a stationary, non-trivial 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 the terms $E\left[\left(C\left(X_{1t}\right) - C\left(X_{2t}\right)\right)\tilde{\theta}_{t+1-j}\right]$, $j = 0, 2, 3, .., p$. This estimation error will lead to reduced power to distinguish between competing RV estimators than would otherwise be the case.

4 Simulation study

To examine the finite-sample performance of the proposed tests I present the results of a small simulation study. I use a standard log-normal stochastic volatility model with a leverage effect, with the same parameters as in Gonçalves and Meddahi (2005):

$$\begin{aligned}
d\log P^*(t) &= 0.0314d(t) + \nu(t)\left(-0.576dW_1(t) + \sqrt{1-0.576^2}dW_2(t)\right) \qquad (6)\\
d\log \nu^2(t) &= -0.0136\left(0.8382 + \log \nu^2(t)\right)d(t) + 0.1148dW_1(t)
\end{aligned}$$

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).

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

$$\begin{aligned}
\log P(t_j) &= \log P^*(t_j) + \xi(t_j) \\
\xi(t_j) &\sim iid\ N\left(0, \sigma_\xi^2\right) \\
\text{where } \frac{2\sigma_\xi^2}{V[r_t]\frac{5}{390} + 2\sigma_\xi^2} &\in \{0, 0.2\}
\end{aligned}$$

where r_t is the open-to-close return on day t. That is, I either set the noise to zero, or I set the variance of the noise 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%. 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). I generated 250 independent sequences of 500 'trade days'.

Existing results on the ARMA processes for integrated variance (IV) implied by various continuous-time stochastic volatility models, see Barndorff-Nielsen and Shephard (2002) and Meddahi (2003), are derived under a 'no leverage' assumption, whereas our simulated process exhibits a leverage effect. This means that the goodness-of-fit of an ARMA model to simulated IV needs to be verified. The results for a selection of ARMA models are presented in Table 1. The first panel of Table 1 shows the average R^2, across simulations, of a random walk model and an AR(1), AR(2), AR(5), ARMA(1,1) and ARMA(2,2). The first point to note is that all of these models fit daily IV very well: The average R^2s ranged from 0.9618 for the random walk model, to 0.9650 for the ARMA(2,2) model. The improvement in average R^2 of the AR(1) model over the random walk model is just 0.0004, while the improvement of the most general model estimated, the ARMA(2,2), over the random walk model is 0.0028. This suggests that although the random walk model is mis-specified for daily IV, it is a reasonable approximation.

The second panel of Table 1 examines in more detail the fit of the AR(1) model for daily IV across simulations. This table shows that the constant in this model is small, at 0.011, but is significantly different from zero: in all 250 simulations the estimate of this parameter is positive. The AR(1) coefficient is near one, averaging 0.981, but the 90^{th} quantile of the cross-simulation distribution of these parameters is 0.991, and this parameter is strictly less than unity in all but three simulations.

[INSERT TABLE 1 ABOUT HERE]

To investigate the finite-sample size and power properties of the tests proposed, I designed the following experiment. For simplicity I focused solely on pair-wise comparisons via a Diebold-Mariano (1995) test. I set the each RV estimator equal to the true IV plus some *iid* noise:

$$
\begin{aligned}
X_{it} &= IV_t + \zeta_{it}, \;\; i = 1, 2 \\
\zeta_{it} &\sim iid\; N\left(0, \sigma_i^2\right) \\
E\left[\zeta_{1t}\zeta_{2s}\right] &= 0 \;\forall\; t, s
\end{aligned}
$$

To study the size of the tests, I first set $\sigma_1^2 = \sigma_2^2 = 0.10 \times V\left[IV_t\right]$. In this case both estimators are noisy estimators of IV, and are equally accurate. Furthermore, in this example both estimators are conditionally unbiased, which means that the satisfy the conditions of Proposition 1 and thus ranking by MSE, using a lead or a lag of the volatility proxy, can be done without any need to bias-correct. For comparison, I also show the results when a bias correction based on an AR(1) assumption is made, following Proposition 5. If the AR(1) is a good approximation to the IV process, then this test should also have reasonable finite-sample properties, though we would expect it to have lower power due to the additional error in the test statistic coming from the estimation of the AR parameters. In estimating the AR(1) parameters I follow Lemma 1, using either zero or three over-identifying moments. To study the power of the tests, I fix σ_1^2 and let $\sigma_2^2/V\left[IV_t\right] = 0.15, 0.20, 0.40, 0.75$. The rejection frequencies under each scenario are presented in Table 2, using both MSE and QLIKE loss.

I consider eleven tests in total. The first two tests are the infeasible Diebold-Mariano (1995) and White (2000) "reality check" tests that one would conduct if the IV were truly observable. The power of these tests represents an upper bound on what we can expect from the feasible tests. I consider the tests under both the random walk assumption (using Proposition 2) and the AR(1) assumption (using Proposition 5 and Lemma 1), for three different proxies: daily squared returns, 30-minute RV and the true IV. The daily squared returns and the 30-minute RV are those obtained from the noisy returns data ($\sigma_\xi^2 > 0$) to examine the impact of microstructure noise on finite-sample size, while the true IV is used to see the limiting case of a proxy with no error being put through these tests.

The first row of each panel of Table 2 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 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 IV were observable; the power is worse when a noisier instrument is used (daily squared returns versus 30-minute RV versus true IV); and the power of the test based on the AR(1) assumption is worse than that based on the random walk assumption. 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. The test based on an AR(1) model estimated using three over-identifying moment conditions generally has slightly better power than the corresponding test based on an estimator with no over-identifying conditions, but the difference in this particular simulation are not large.

In addition to testing the size and power of tests based on the results of the previous section, it is also of interest to study the simple rankings of estimators obtained using the methods proposed in this paper. Table 3 reports the proportion of simulations for which the ranking of the two competing estimators obtained using one of the methods proposed above is the same as the ranking that would be obtained using the true IV. In the first row of each panel the two RV estimators are equally accurate, and so we expect the proportion of correct rankings to lie anywhere between zero and one. As we move down the rows in each column, we expect the proportion of correct rankings to increase towards unity, with both methods correctly identifying the better estimator. This is indeed what is found. The proportion of correct rankings increases as the difference in accuracy of the two estimators increases; decreases as the noise in the volatility proxy increases; and weakly decreases as we move from the random walk to the AR(1) assumption. No differences are found between rankings based on an AR(1) model estimated using three over-identifying moment conditions versus no over-identifying conditions.

Overall, I conclude that the proposed tests have reasonable finite-sample size properties. There is of course a loss in power when using a noisy proxy rather than the true, unobservable, target variable, however this is the nature of the problem under analysis. Table 3 shows that although there is a loss in power, the rankings obtained using the proposed methods generally consistent with the rankings based on the unobservable target variable.

[INSERT TABLES 2 AND 3 ABOUT HERE]

5 Empirical application

In this section I consider the problem of estimating the quadratic variation of the daily return on IBM. I use data on trade prices from the TAQ database over the period from January 1993 to May 1998, yielding 1364 daily returns. This sample period was used in the Andersen, *et al.* (2001) study of realised variance of equity returns. I consider two types of simple realised volatility estimators: the first using "calendar time" sampling, and the second using "tick time" sampling. The existing literature provides little guidance on calendar-time versus tick-time sampling. Oomen (2006) is a notable exception to this, and via a parametric pure jump process for transaction prices he finds that tick-time sampling has lower MSE than calendar-time sampling.

For the first type of RV estimator, I use "last price" interpolation of the trade price series to create a series of 30-second prices for the open hours of the New York Stock exchange (9:30am to 4:00pm), denoted $\{r_{t,j},\ j = 1, 2, ..., 780\}_{t=1}^{T}$. From this series I compute:

$$\begin{aligned} RV_t^{(m)} &= \sum_{j=1}^{m} r_{t,j,m}^2 \\ \text{where} \quad r_{t,j,m} &\equiv \sum_{i=1}^{2h} r_{t,2h(j-1)+i} \\ h &= \frac{390}{m} \end{aligned}$$

$r_{t,j,m}$ is the h-minute return, computed from the original 30-second return series. I consider all values for h that are multiples of one-half (so that I can evenly aggregate these from the original 30-second return series) and that divide evenly into 390, the number of minutes in an NYSE trade day. This yields 17 sampling frequencies: $h = 0.5$, 1, 2, 3, 5, 6, 10, 13, 15, 26, 30, 39, 65, 78, 130, 195, and 390 minutes, the final value for h corresponding to simply using the open-to-close return. The corresponding values for m are 780, 390, 195, 130, 78, 65, 39, 30, 26, 13, 10, 6, 5, 3, 2, 1.

The "tick time" RV estimators, denoted $RVtick_t^{(m)}$, are constructed by sampling the trade prices so that I have m return observations per day: for example, for $m = 2$ I sample the first trade, last trade and the trade closest to the middle trade (in terms of number of trades, not in terms of clock time) and use these to compute the 2 return observations. I use the same set of values for m that are used for the standard RV estimators.

The total number of RV estimators considered is 33: 17 standard RV estimators and 16 RVtick estimators (for $m = 1$, the calendar-time and tick-time estimators are identical and so I drop the last *RVtick* estimator). Figure 1 presents the volatility signature plot and a plot of the standard deviation of these estimators.

[INSERT FIGURE 1 ABOUT HERE]

Tables 4 to 6 present the first empirical contribution of this paper. These tables present the average distance, under MSE and QLIKE, between the 33 RV estimators and the latent target variable relative to the average distance between the squared open-to-close ("daily") return and the latent target variable. A negative value indicates that the daily squared return was out-performed, while a positive value indicates the opposite. In all cases the proxy is the squared daily return. The three tables show the estimated average distances under three assumptions on the DGP: Table 4 is based on an AR(1) assumption for the latent target variable (Assumption T2") and computes the average distance differentials using the consistent estimator presented in Proposition 5, using three extra moments to estimate the AR(1) parameters. Table 5 is based on a random walk assumption for the latent target variable (Assumption T2) and computes the average distance differentials using the estimator presented in Proposition 2. Table 6 is based on the incorrect assumption that the measurement error in the proxy is uncorrelated with the errors in the RV estimators, and uses the contemporaneous value of the proxy rather than a one-period lead as in the former two cases. The results from these three tables are depicted in Figure 2.

[INSERT FIGURE 2 ABOUT HERE]

Under the AR(1) assumption for the target variable, the best two estimators according to MSE and QLIKE are the RV estimators based on 30-minute and 15-minute returns. The worst two estimators under MSE are the RV-tick estimators based on $m = 780$ and $m = 390$ trades (corresponding to $h = 0.5$ minute and $h = 1$ minute sampling on average). The worst estimators under QLIKE are daily squared returns and RV based on 195-minute returns.

Under the random walk assumption for the target variable, the best RV estimators according to MSE and QLIKE are the RV estimators based on 30-minute and $m = 78$ trade (2-minute) returns respectively. The second-best estimators are those based on h =5-minute returns and m =130 trades. The worst estimators are the same as those under the AR(1) assumption.

To illustrate the distortions caused by ignoring the correlation between the error in the proxy and the RV estimators, I also present the ranking obtained under the naïve assumption that this correlation is zero. The resulting ranking suggests that daily squared returns are the *best* estimator of daily quadratic variation amongst all RV and RV-tick estimators, which is driven purely by the fact that the correlation between the measurement errors goes to unity for the standard RV estimator when $m = 1$; far from the assumption that it is zero.

[INSERT TABLES 4, 5 AND 6 ABOUT HERE]

In Table 7 I present the results of formal comparisons of the 33 RV estimators considered in this empirical application. To do this I implement the "reality check" of White (2000), and a refinement of this test proposed by Hansen (2005). The reality check is a means of testing the null:

$$\begin{aligned} H_0 &: E\left[L\left(\theta_t, X_{0t}\right)\right] \leq E\left[L\left(\theta_t, X_{it}\right)\right], \text{ for all } i = 1, 2, ..., K \\ \text{vs. } H_a &: E\left[L\left(\theta_t, X_{0t}\right)\right] > E\left[L\left(\theta_t, X_{it}\right)\right] \text{ for some } i \end{aligned}$$

where X_{0t} is some benchmark RV estimator. That is, we test whether the benchmark RV estimator generates losses that are weakly smaller in expectation than any competing RV estimator. The null hypothesis contains K weak inequalities, and the critical values for this test can be easily obtained using a bootstrap procedure. I use the stationary bootstrap of Politis and Romano (1994) with an average block length of 20 days. Using the bootstrap also simplifies accounting for the impact of the estimator of the AR(1) parameters on the asymptotic distribution of the test statistic, see Corollary 2.7 of White (2000). Hansen's (2005) refinement of the White's reality check involves a form of "trimming" to limit the impact of very poor estimators and studentising the test statistic; both of these refinements should lead to improved power to reject the null.

I consider four benchmark estimators of daily quadratic variation: the daily squared return, a standard RV estimator based on 5-minute returns, an RV-tick estimator based on 78 trades per day (5-minute returns on average) and the estimated volatility obtained from a Normal GARCH(1,1) model applied to the open-to-close return series. I present results under both the AR(1) assumption and the random walk assumption, which allows for some insight into the impact of estimation error in the AR(1) parameter estimate on the power of the test. Finally, I consider two proxies: the squared daily return, and a standard RV estimator based on 2intra-daily returns $(h = 195)$. This latter estimator is approximately unbiased and is about 30% less volatile than daily squared returns, according to the plots in Figure 1, and so may lead to more powerful inference.

[INSERT TABLE 7 ABOUT HERE]

Table 7 reveals that the daily squared return can be rejected as being significantly beaten by some alternative RV estimator in many cases: for all but one case under the random walk assumption it is rejected, as well as under the AR(1) assumption and the QLIKE pseudo-distance measure. When using the AR(1) assumption the daily squared return is mostly not rejected either under MSE or QLIKE, perhaps indicating a loss of power for this application.

The standard RV estimator based on 5-minute returns is rejected only twice, suggesting that for this sample period the competing RV or RV-tick estimators were not generally significantly better than this simple estimator. Similarly, the RV-tick estimate based on 78 returns per day is not rejected by any test. This finding provides some support for the rule-of-thumb that a simple 5-minute RV estimator (either in calendar time or in tick time) works well in practice.

For comparison, I also considered the estimated volatility from a simple GARCH(1,1) model (see Engle, 1982, and Bollerslev, 1986) as a measure of daily quadratic variation. This estimator is almost certainly biased relative to RV estimators based on the current day's information, as the GARCH estimate for day t uses only data up until day $t-1$, however the GARCH estimates will be smoother than the RV estimates, perhaps allowing for some bias-variance trade-off. This is indeed what is found: in no case is the GARCH estimator rejected in favour of one of the competing RV or RV-tick estimators.

Overall, this small empirical application suggests that it is difficult to beat simple estimates of daily quadratic variation. Daily squared returns are significantly beaten by estimators that use intra-daily data, but a standard RV estimator based on 5-minute returns (computed either in calendar time or in tick time), and even an estimate obtained from a GARCH(1,1) model are not significantly out-performed by estimators based on higher frequency data. It remains to be seen whether this conclusion holds for other assets in other sample periods.

6 Conclusion

This paper considers the problem of comparing realised volatility (RV) estimators. I propose a data-based method for formally ranking RV estimators that does not rely on simulations, detailed assumptions about the market microstructure noise process, or on "large m" (or "continuous record") asymptotics, though my method does rely on "large T" asymptotics. By either imposing some assumptions on the time series dynamics of the biases in the RV estimators, or by imposing a rather weak assumption on the time series dynamics of the latent target variable, I present results that allow for a consistent estimate of the ranking of competing RV estimators. These results can be used in formal Diebold-Mariano (1995) pair-wise comparisons of RV estimators, or comparisons involving multiple estimators, such as the "reality check" of White (2000) or its refinement by Hansen (2005).

In a small empirical application to IBM equity return volatility, I find evidence that the daily squared return is out-performed as a measure of quadratic variation by RV estimators based on higher frequency data. However, I find little evidence that a simple RV estimator constructed using 5-minute returns (either in calendar time or in "tick time") is out-performed by estimators using higher frequency data.

This paper immediately suggests two extensions, which are being pursued in separate work. The first is the comparison of different realised covariance estimators. The methods presented in this paper apply directly to this case, subject to a suitable pseudo-distance measure being selected in place of the parametric family presented in equation (2). The standard squared difference distance measure (MSE) is applicable for comparing realised covariance estimators, and other measures for this case are discussed in Patton (2006). The second important 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 assumption for the latent integrated covariance matrix.

7 Appendix: Proofs

Proof of Proposition 1. Consider a first-order Taylor series expansion of $C(X_t; b)$

$$\begin{aligned}
C(X_t; b) &\approx C(\theta_t; b) + C'(\theta_t; b)(X_t - \theta_t) \\
\text{so } E[(C(X_t; b) - C(\theta_t; b))(Y_t - \theta_t)] &\approx E\left[C'(\theta_t; b)(X_t - \theta_t)(Y_t - \theta_t)\right] \\
&= -E\left[\theta_t^{-k}(X_t - \theta_t)(Y_t - \theta_t)\right]
\end{aligned}$$

Under assumption P2 we have:

$$\begin{aligned}
E\left[\theta_t^{-k}(X_t - \theta_t)(Y_t - \theta_t)\right] &\approx E\left[\theta_t^{-k}(E[X_t|\theta_t, \mathcal{F}_{t-1}] - \theta_t)(Y_t - \theta_t)\right] \\
&= E\left[\theta_t^{-k}\left(c_i\theta_t^k\right)(Y_t - \theta_t)\right] \\
&= c_i E[Y_t - \theta_t] \\
&= 0
\end{aligned}$$

Thus we have

$$\begin{aligned}
E[L(Y_t, X_{1t}; b)] &= E[L(\theta_t, X_t; b)] + E[(C(X_t; b) - C(\theta_t; b))(Y_t - \theta_t)] \\
&\quad -\frac{1}{2}E\left[C'\left(\ddot{\theta}_t; b\right)(Y_t - \theta_t)^2\right] \\
&\approx E[L(\theta_t, X_t; b)] - \frac{1}{2}E\left[C'\left(\ddot{\theta}_t; b\right)(Y_t - \theta_t)^2\right] \\
\text{and so } E[L(Y_t, X_{1t}; b)] - E[L(Y_t, X_{2t}; b)] &\approx E[L(\theta_t, X_{1t}; b)] - E[L(\theta_t, X_{2t}; b)]
\end{aligned}$$

up to the error term from the first-order Taylor series expansion of $C(X_t; b)$ around $C(\theta_t; b)$. When $b = 0$, $C(z; b) = -z$ and so the first-order Taylor series expansion holds exactly. Furthermore, when $b = 0$ this distance measure can be applied to both positive and negative variables. ■

Proof of Proposition 2. Consider the expectation second term in the second-order mean-value expansion of $L(Y_t, X_t; b)$ around $L(\theta_t, X_t; b)$:

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

This then yields

$$E\left[L\left(Y_t,X_{1t};b\right)\right]-E\left[L\left(Y_t,X_{2t};b\right)\right]=E\left[L\left(\theta_t,X_{1t};b\right)\right]-E\left[L\left(\theta_t,X_{2t};b\right)\right]$$

using the same calculations as in the proof of Proposition 1. ■

Proof of Proposition 3. (i) Consider again the expectation second term in the mean-value expansion of $L(Y_t, X_t; b)$ around $L(\theta_t, X_t; b)$:

$$\begin{aligned}
E\left[\left(C\left(X_t;b\right)-C\left(\theta_t;b\right)\right)\left(Y_t-\theta_t\right)\right] &= E\left[\left(C\left(X_t;b\right)-C\left(\theta_t;b\right)\right)\left(\tilde{\theta}_{t+1}-\theta_t\right)\right] \\
&= E\left[\left(C\left(X_t;b\right)-C\left(\theta_t;b\right)\right)\left(\bar{\theta}\delta^2-\delta\theta_t+\eta_{t+1}+\nu_{t+1}\right)\right] \\
&= \bar{\theta}\delta^2E\left[C\left(X_t;b\right)-C\left(\theta_t;b\right)\right]-\delta E\left[\theta_t\left(C\left(X_t;b\right)-C\left(\theta_t;b\right)\right)\right]
\end{aligned}$$

(ii) First consider the case that $b=0$:

$$\begin{aligned}
E\left[\theta_t\left(C\left(X_{1t};b\right)-C\left(X_{2t};b\right)\right)\right] &= -E\left[\theta_t\left(X_{1t}-X_{2t}\right)\right] \\
&= -E\left[\theta_t^{k+1}\left(c_1-c_2\right)+\theta_t\left(\xi_{1t}-\xi_{2t}\right)\right] \\
&= -\epsilon E\left[\theta_t^{k+1}\right] \\
\text{since } \; E\left[\xi_{it}|\theta_t,\mathcal{F}_{t-1}\right] &= 0,\; i=1,2
\end{aligned}$$

For cases with $b \neq 0$ I employ a first-order Taylor series approximation of $C(X_{it}; b)$ around $C(\theta_t; b)$

$$\begin{aligned}
C\left(X_{it};b\right) &\approx C\left(\theta_t;b\right)+C'\left(\theta_t;b\right)\left(X_{it}-\theta_t\right) \\
\text{so } \; E\left[\theta_t\left(C\left(X_{1t};b\right)-C\left(X_{2t};b\right)\right)\right] &\approx E\left[\theta_t\left(-\theta_t^b\left(X_{1t}-\theta_t\right)+\theta_t^b\left(X_{2t}-\theta_t\right)\right)\right] \\
&= -E\left[\theta_t^{b+1}\left(X_{1t}-X_{2t}\right)\right] \\
&= -\epsilon E\left[\theta_t^{k+1+b}\right]
\end{aligned}$$

■

Proof of Proposition 4.

$$
\begin{aligned}
Y_t &= \frac{1}{J}\sum_{i=1}^{J}\left(\theta_{t+i}+\nu_{t+i}\right) \\
&= \frac{1}{J}\sum_{i=1}^{J}\left(\nu_{t+i}+\theta_t+\sum_{k=1}^{i}\eta_{t+k}\right) \\
\text{so } Y_t-\theta_t &= \frac{1}{J}\sum_{i=1}^{J}\nu_{t+i}+\frac{1}{J}\sum_{i=1}^{J}\sum_{k=1}^{i}\eta_{t+k} \\
&= \frac{1}{J}\sum_{i=1}^{J}\nu_{t+i}+\frac{1}{J}\sum_{i=1}^{J}\left(J+1-i\right)\eta_{t+i}
\end{aligned}
$$

From this expression I compute the MSE of Y_t, as a function of J and $\sigma_\eta^2 \equiv E\left[\eta_t^2\right]$, $\sigma_\nu^2 \equiv E\left[\nu_t^2\right]$ and $\sigma_{\eta\nu} \equiv E\left[\eta_t \nu_t\right]$

$$
\begin{aligned}
E\left[\left(Y_t-\theta_t\right)^2\right] &= \frac{1}{J^2}E\left[\left(\sum_{i=1}^{J}\nu_{t+i}\right)^2\right]+\frac{1}{J^2}E\left[\left(\sum_{i=1}^{J}\left(J+1-i\right)\eta_{t+i}\right)^2\right] \\
&\quad +\frac{2}{J^2}E\left[\left(\sum_{i=1}^{J}\nu_{t+i}\right)\left(\sum_{i=1}^{J}\left(J+1-i\right)\eta_{t+i}\right)\right] \\
&= \frac{1}{J}\sigma_\nu^2+\frac{1}{J^2}\sigma_\eta^2\sum_{i=1}^{J}\left(J+1-i\right)^2+\frac{2}{J^2}\sigma_{\eta\nu}\sum_{i=1}^{J}\left(J+1-i\right) \\
&= \frac{1}{J}\sigma_\nu^2+\frac{\left(J+1\right)\left(2J+1\right)}{6J}\sigma_\eta^2+\left(1+\frac{1}{J}\right)\sigma_{\eta\nu}
\end{aligned}
$$

This expression reveals the competing influences of the three terms: the first term is decreasing in J, the second term is increasing in J, and the third term is approximately flat in J for large values of J; for small J it is decreasing in J.

w.l.o.g., let $\sigma_\nu^2 = \psi\sigma_\eta^2$ and so $\sigma_{\eta\nu} \equiv \rho\sqrt{\psi}\sigma_\eta^2$. In that case the first-order condition for J^* becomes:

$$
\begin{aligned}
0 &= -\frac{\sigma_\eta^2}{6J^2}\left(-1+2J^{*2}-6\psi-6\rho\sqrt{\psi}\right) \\
\text{so } J^* &= \sqrt{\frac{1+6\psi+6\rho\sqrt{\psi}}{2}}
\end{aligned}
$$

The optimal value of J for various values of ψ and ρ is given below:

J^*	ρ				
ψ	-0.9	-0.5	0	0.5	0.9
0.0001	0.69	0.70	0.71	0.72	0.73
0.1	n/a	0.57	0.89	1.13	1.29
1	0.89	1.41	1.87	2.24	2.49
10	4.69	5.08	5.52	5.94	6.25
100	16.54	16.90	17.33	17.76	18.10
10,000	172.43	172.77	173.21	173.64	173.98

When we constrain J^* to be an integer between 1 and 10000, the optimal values are those presented in the statement of the proposition. ∎

Proof of Lemma 1. Recall that autocovariances of an AR(p) process satisfy:

$$\begin{aligned} \gamma_j &\equiv Cov\left[\theta_t, \theta_{t-j}\right] = \sum_{i=1}^{p} \phi_i \gamma_{j-i} \text{ for } j \geq p \\ \text{and that } Cov\left[\tilde{\theta}_t, \tilde{\theta}_{t-j}\right] &= Cov\left[\theta_t, \theta_{t-j}\right] \ \forall\ j \neq 0 \end{aligned}$$

under assumption P1'. Thus

$$A^{(k)}\tilde{\Phi} = B^{(k)} \ \forall\ k \geq 0$$

when $\tilde{\Phi} = \Phi$, the true vector of AR(p) parameters. This suggests the general minimum distance estimator defined by

$$\hat{\Phi}_T \equiv \arg\min_{\Phi} \left(\hat{A}_T^{(k)\prime}\Phi - \hat{B}_T^{(k)}\right)' W_T \left(\hat{A}_T^{(k)\prime}\Phi - \hat{B}_T^{(k)}\right)$$

which has first-order condition

$$\begin{aligned} 0 &= \hat{A}_T^{(k)\prime} W_T \left(\hat{A}_T^{(k)}\hat{\Phi}_T - \hat{B}_T^{(k)}\right) \\ \text{so } \hat{\Phi}_T &= \left(\hat{A}_T^{(k)\prime} W_T \hat{A}_T^{(k)}\right)^{-1} \hat{A}_T^{(k)\prime} W_T \hat{B}_T^{(k)} \end{aligned}$$

where $\hat{A}_T^{(k)\prime} W_T \hat{A}_T^{(k)\prime}$ is positive definite if W_T is positive definite. For the asymptotic covariance matrix of this estimator see Baillie and Chung (2001). ∎

Proof of Proposition 5. (i) Recall the second-order mean-value expansion of the loss function from the proof of Proposition 1, leading to

$$\begin{aligned} E\left[L\left(Y_t, X_{1t}; b\right)\right] - E\left[L\left(Y_t, X_{2t}; b\right)\right] &= E\left[L\left(\theta_t, X_{1t}; b\right)\right] - E\left[L\left(\theta_t, X_{2t}; b\right)\right] \\ &\quad + E\left[\left(C\left(X_{1t}; b\right) - C\left(X_{2t}; b\right)\right)\left(Y_t - \theta_t\right)\right] \end{aligned}$$

If $Y_t = \tilde{\theta}_{t+1}$ and assumption T2" is satisfied, set $\phi_0 \equiv \mu\left(1 - \sum_{i=1}^{p} \phi_i\right)$ and then

$$\begin{aligned} E\left[\left(C\left(X_{1t}; b\right) - C\left(X_{2t}; b\right)\right) Y_t\right] &= E\left[\left(C\left(X_{1t}; b\right) - C\left(X_{2t}; b\right)\right)\left(\theta_{t+1} + \nu_{t+1}\right)\right] \\ &= E\left[\left(C\left(X_{1t}; b\right) - C\left(X_{2t}; b\right)\right) \theta_{t+1}\right] \\ &= E\left[\left(C\left(X_{1t}; b\right) - C\left(X_{2t}; b\right)\right) E_t\left[\theta_{t+1}\right]\right] \\ &= \phi_0 E\left[C\left(X_{1t}; b\right) - C\left(X_{2t}; b\right)\right] \\ &\quad + \phi_1 E\left[\left(C\left(X_{1t}; b\right) - C\left(X_{2t}; b\right)\right) \theta_t\right] \\ &\quad + \sum_{i=2}^{p} \phi_i E\left[\left(C\left(X_{1t}; b\right) - C\left(X_{2t}; b\right)\right) \theta_{t+1-i}\right] \\ \text{so} \quad E\left[\left(C\left(X_{1t}; b\right) - C\left(X_{2t}; b\right)\right) \theta_t\right] &= \frac{1}{\phi_1} E\left[\left(C\left(X_{1t}; b\right) - C\left(X_{2t}; b\right)\right) \tilde{\theta}_{t+1}\right] \\ &\quad - \frac{\phi_0}{\phi_1} E\left[C\left(X_{1t}; b\right) - C\left(X_{2t}; b\right)\right] \\ &\quad - \sum_{i=2}^{p} \frac{\phi_i}{\phi_1} E\left[\left(C\left(X_{1t}; b\right) - C\left(X_{2t}; b\right)\right) \theta_{t+1-i}\right] \end{aligned}$$

Next note that

$$\begin{aligned} E\left[\left(C\left(X_{1t}; b\right) - C\left(X_{2t}; b\right)\right) \tilde{\theta}_{t+1-i}\right] &= E\left[\left(C\left(X_{1t}; b\right) - C\left(X_{2t}; b\right)\right)\left(\theta_{t+1-i} + \nu_{t+1-i}\right)\right] \\ &= E\left[\left(C\left(X_{1t}; b\right) - C\left(X_{2t}; b\right)\right) \theta_{t+1-i}\right] \ \forall\ i \geq 2 \end{aligned}$$

under assumption R2. So

$$\begin{aligned} E\left[\left(C\left(X_{1t}; b\right) - C\left(X_{2t}; b\right)\right)\left(Y_t - \theta_t\right)\right] &= \frac{\phi_0}{\phi_1} E\left[C\left(X_{1t}; b\right) - C\left(X_{2t}; b\right)\right] \\ &\quad + \sum_{i=2}^{p} \frac{\phi_i}{\phi_1} E\left[\left(C\left(X_{1t}; b\right) - C\left(X_{2t}; b\right)\right) \tilde{\theta}_{t+1-i}\right] \\ &\quad - \frac{1 - \phi_1}{\phi_1} E\left[\left(C\left(X_{1t}; b\right) - C\left(X_{2t}; b\right)\right) \tilde{\theta}_{t+1}\right] \end{aligned}$$

and thus

$$\begin{aligned} E\left[L\left(\theta_t, X_{1t}; b\right)\right] - E\left[L\left(\theta_t, X_{2t}; b\right)\right] &= E\left[L\left(Y_t, X_{1t}; b\right)\right] - E\left[L\left(Y_t, X_{2t}; b\right)\right] \\ &\quad - \frac{\phi_0}{\phi_1} E\left[C\left(X_{1t}; b\right) - C\left(X_{2t}; b\right)\right] \\ &\quad - \sum_{i=2}^{p} \frac{\phi_i}{\phi_1} E\left[\left(C\left(X_{1t}; b\right) - C\left(X_{2t}; b\right)\right) \tilde{\theta}_{t+1-i}\right] \\ &\quad + \frac{1-\phi_1}{\phi_1} E\left[\left(C\left(X_{1t}; b\right) - C\left(X_{2t}; b\right)\right) \tilde{\theta}_{t+1}\right] \end{aligned}$$

(ii) To implement this method, we need consistent estimators of $\left(\phi_0, \phi_1, ..., \phi_p\right)$,which are obtained using Lemma 1. ■

Table 1a: Goodness-of-fit of some ARMA models for daily Integrated Variance

	Random walk	AR(1)	AR(2)	AR(5)	ARMA(1,1)	ARMA(2,2)
Average R^2	0.9618	0.9622	0.9627	0.9631	0.9648	0.9650

.

Table 1b: Properties of the AR(1) model for Integrated Variance across simulations

	Mean	5% quantile	95% quantile
Constant	0.011	0.003	0.021
AR(1) coeff	0.981	0.967	0.991
R^2	0.962	0.935	0.979

Notes: The first panel of this table presents the average R^2 coefficients, across 250 simulations, for six different models for the daily integrated variance resulting from the continuous time stochastic volatility model described in equation (6). The second panel presents the average, and the 5% and 95% quantiles, of the estimated coefficients of the AR(1) model for daily integrated variance, and the associated R^2, across 200 simulations.

Table 2: Finite-sample size and power of the tests

Proxy	*IV*		*IV*			*RV-30min*			*RV Daily*		
	Diebold-Mariano	Reality Check	Random walk	AR(1) k=0	AR(1) k=3	Random walk	AR(1) k=0	AR(1) k=3	Random walk	AR(1) k=0	AR(1) k=3
$\sigma_2^2/V\left[IV_t\right]$											
						MSE					
0.10	0.059	0.005	0.022	0.005	0.011	0.032	0.022	0.022	0.032	0.038	0.032
0.15	0.995	0.935	0.897	0.822	0.838	0.130	0.103	0.103	0.108	0.027	0.038
0.20	1.000	1.000	1.000	0.995	1.000	0.211	0.162	0.216	0.114	0.032	0.049
0.40	1.000	1.000	1.000	1.000	1.000	0.481	0.481	0.476	0.297	0.103	0.146
0.75	1.000	1.000	1.000	1.000	1.000	0.595	0.584	0.605	0.459	0.141	0.227
						QLIKE					
0.10	0.016	0.000	0.000	0.005	0.000	0.022	0.016	0.011	0.032	0.054	0.011
0.15	0.297	0.076	0.070	0.076	0.103	0.054	0.054	0.059	0.032	0.119	0.032
0.20	0.427	0.184	0.162	0.249	0.216	0.076	0.086	0.054	0.043	0.092	0.032
0.40	0.562	0.432	0.341	0.449	0.389	0.232	0.216	0.200	0.092	0.141	0.038
0.75	0.611	0.492	0.535	0.551	0.568	0.307	0.276	0.319	0.151	0.119	0.032

Notes: This table presents the rejection frequencies for eight different tests of equal accuracy of two competing RV estimators. All tests are conducted at the 0.05 level, and 250 simulations are used. 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 (σ_2^2) than the first $(\sigma_1^2 = 0.10 \times V\left[IV_t\right])$. For the tests based on the AR(1) assumption, k represents the number of over-identifying conditions used in estimation.

Table 3: Proportion of correct rankings of estimators

Proxy	*IV*			*RV-30min*			*RV Daily*		
	Random	AR(1)		Random	AR(1)		Random	AR(1)	
$\sigma_2^2/V\,[IV_t]$	walk	k=0	k=3	walk	k=0	k=3	walk	k=0	k=3
				MSE					
0.10	0.81	0.80	0.80	0.65	0.66	0.66	0.56	0.58	0.58
0.15	1.00	1.00	1.00	0.92	0.92	0.92	0.66	0.58	0.58
0.20	1.00	1.00	1.00	0.98	0.98	0.98	0.76	0.72	0.72
0.40	1.00	1.00	1.00	1.00	1.00	1.00	0.95	0.89	0.89
0.75	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.92	0.92
				QLIKE					
0.10	0.96	0.96	0.96	0.84	0.83	0.83	0.64	0.51	0.51
0.15	0.97	0.98	0.98	0.92	0.91	0.91	0.75	0.61	0.61
0.20	1.00	1.00	1.00	0.97	0.96	0.96	0.78	0.68	0.68
0.40	0.99	0.99	0.99	0.97	0.97	0.97	0.88	0.72	0.72
0.75	1.00	0.99	0.99	0.99	0.99	0.99	0.92	0.81	0.81

Notes: This table presents the proportion of simulations in which the ranking of two competing RV estimators, obtained using one of the data-based methods proposed in this paper, is the same as that obtained using the true IV. 250 simulations are used. The competing RV estimators are equally accurate on average in the first row of each panel, while in the other rows the second RV estimator has greater noise variance (σ_2^2) than the first ($\sigma_1^2 = 0.10 \times V\,[IV_t]$). For the tests based on the AR(1) assumption, k represents the number of over-identifying conditions used in estimation.

Table 4: Average distance relative to daily squared returns, under the AR(1) assumption

AR(1)				
	MSE		*QLIKE*	
h	RV	RV-tick	RV	RV-tick
0.5	-7.23	81.68	-15.28	-14.73
1	-10.05	7.13	-15.32	-15.06
2	-11.75	-9.07	-15.37	-15.28
3	-11.50	-10.33	-15.39	-15.32
5	-12.93	-11.99	-15.39	*-15.40*
6	-11.73	-11.08	-15.34	-15.38
10	-12.82	-11.73	-15.35	-15.36
13	-12.05	-10.71	-15.33	-15.23
15	-13.03	-11.09	**-15.42**	-15.36
26	*-13.20*	-11.17	-15.23	-15.25
30	**-14.44**	-12.11	-15.13	-15.26
39	-10.99	-11.81	-14.95	-15.20
65	-10.04	-10.81	-14.45	-15.24
78	-9.95	-12.54	-14.34	-14.81
130	-10.50	-7.88	-13.51	-13.43
195	-9.85	-8.97	-10.13	-12.19
390	0.00	n/a	0.00	n/a

Notes: This table presents the estimated difference in the mean distance between the latent integrated variance and an RV estimator based on h-minute or h-tick sampling, and the mean distance the latent integrated variance and the squared open-to-close return, using the bias-adjustment from Proposition 5 with $k = 3$. The best forecast for a given pseudo-distance measure is in bold; the second-best is in italics.

Table 5: Average distance relative to daily squared returns, under the random walk assumption

RW				
	MSE		*QLIKE*	
h	RV	RV-tick	RV	RV-tick
0.5	-9.85	86.63	-18.42	-18.00
1	-14.10	6.28	-18.45	-18.29
2	-16.66	-11.72	-18.46	-18.44
3	-16.61	-14.48	-18.45	*-18.46*
5	*-17.13*	-16.40	-18.41	**-18.48**
6	-16.53	-15.59	-18.37	-18.45
10	-16.75	-16.43	-18.31	-18.40
13	-16.73	-15.88	-18.27	-18.30
15	-17.01	-15.98	-18.28	-18.34
26	-16.81	-15.64	-18.07	-18.20
30	**-17.36**	-16.30	-17.98	-18.18
39	-14.89	-16.03	-17.82	-18.08
65	-13.57	-13.93	-17.16	-17.86
78	-12.78	-15.37	-16.94	-17.59
130	-12.98	-10.66	-15.37	-16.19
195	-10.83	-9.84	-12.18	-14.08
390	0.00	n/a	n/a	0.00

Notes: This table presents the estimated difference in the mean distance between the latent integrated variance and an RV estimator based on h-minute or h-tick sampling, and the mean distance the latent integrated variance and the squared open-to-close return, assuming that the integrated variance follows a random walk. The best forecast for a given pseudo-distance measure is in bold; the second-best is in italics.

Table 6: Average distance relative to daily squared returns, ignoring correlation in measurement errors

Naïve				
	MSE		*QLIKE*	
h	RV	RV-tick	RV	RV-tick
0.5	19.42	103.61	1.09	1.65
1	17.42	34.24	1.04	1.30
2	16.53	19.39	0.98	1.09
3	16.06	18.59	0.96	1.03
5	16.15	16.98	0.95	0.99
6	15.15	17.19	0.94	0.98
10	14.94	15.55	0.93	0.96
13	15.78	16.35	0.93	0.94
15	15.46	15.37	0.93	0.95
26	14.43	15.77	0.91	0.92
30	15.36	15.78	0.91	0.91
39	12.95	15.32	0.88	0.91
65	11.84	12.32	*0.85*	0.87
78	10.49	13.49	0.86	0.88
130	9.76	9.97	1.08	0.88
195	*7.52*	8.76	0.88	0.89
390	**0.00**	n/a	**0.00**	n/a

Notes: This table presents the estimated difference in the mean distance between the latent integrated variance and an RV estimator based on h-minute or h-tick sampling, and the mean distance the latent integrated variance and the squared open-to-close return, ignoring the correlation between the measurement error in the proxy and the error in the RV estimators. The best forecast for a given pseudo-distance measure is in bold; the second-best is in italics.

Table 7: P-values from "reality check" tests

			MSE		*QLIKE*	
Benchmark	DGP	Proxy	White	Hansen	White	Hansen
Daily	AR(1)	Daily	0.270	0.355	**0.064**	0.425
Daily	AR(1)	195-min RV	0.328	0.448	**0.032**	0.481
Daily	RW	Daily	0.227	**0.013**	**0.001**	**0.001**
Daily	RW	195-min RV	**0.008**	**0.007**	**0.000**	**0.000**
5-min RV	AR(1)	Daily	0.895	0.576	1.000	0.949
5-min RV	AR(1)	195-min RV	0.872	0.621	1.000	0.940
5-min RV	RW	Daily	0.984	0.583	0.954	**0.014**
5-min RV	RW	195-min RV	0.961	0.484	0.988	**0.028**
5-min RV-tick	AR(1)	Daily	0.833	0.792	1.000	0.784
5-min RV-tick	AR(1)	195-min RV	0.755	0.676	0.984	0.748
5-min RV-tick	RW	Daily	0.918	0.403	1.000	0.791
5-min RV-tick	RW	195-min RV	0.836	0.204	1.000	0.720
GARCH	AR(1)	Daily	0.643	0.621	0.869	0.628
GARCH	AR(1)	195-min RV	0.672	0.597	0.776	0.620
GARCH	RW	Daily	0.828	0.312	0.981	0.426
GARCH	RW	195-min RV	0.872	0.376	0.971	0.400

Notes: This table presents the p-values from the reality check of White (2000), and those from Hansen's (2005) refinement of the reality check, under MSE and QLIKE pseudo-distance measures. Two assumptions on the DGP for the target variable: the AR(1) assumption corresponds to Assumption T2", while the RW assumption corresponds to T2. Four benchmark forecasts are considered: RV based on $h = 390$ (denoted "daily"), RV based on 5 minutes returns, RV based on 5 minute average "ticks", and the estimated volatility produced by a GARCH(1,1) model estimated on the full sample of open-to-close returns. Two variables were considered as the proxy: the squared open-to-close return (denoted "daily") and standard realised variance based on 2 returns per day (denoted "195-min RV"). In both cases the instrument was the one-period lead of the proxy variable. A p-value of less than 0.10 indicates that the benchmark RV estimator is significantly beaten by one of the competing RV estimators at the 10% level.

References

[1] Aït-Sahalia, Yacine, Mykland, Per and Zhang, Lan, 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, Torben G., Bollerslev, Tim, and Meddahi, Nour, 2004, Analytic Evaluation of Volatility Forecasts, *International Economic Review*, 45, 1079-1110.

[3] Andersen, Torben G., Bollerslev, Tim, Diebold, Francis X. and Labys, Paul, 2000, Great Realizations, *Risk*, 13, 105-108.

[4] Andersen, Torben G., Bollerslev, Tim, Christoffersen, Peter F., and Diebold, Francis X., 2005, Volatility and Correlation Forecasting, in G. Elliott, A. Timmermann and C.W.J. Granger (ed.s), *Handbook of Economic Forecasting*, North Holland, forthcoming.

[5] Andersen, Torben, Bollerslev, Tim, Diebold, Francis X., and Ebens, Heiko, 2001, The Distribution of Realized Stock Return Volatility, *Journal of Financial Economics*, 61, 43-76.

[6] Baillie, Richard T., and Chung, Humin, 2001, Estimation of GARCH Models from the Autocorrelations of the Squares of a Process, *Journal of Time Series Analysis*, 22(6), 631-650.

[7] Bandi, Federico M., and Russell, Jeffrey R., 2005, Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations, working paper, Graduate School of Business, University of Chicago.

[8] Barndorff-Nielsen, Ole E., and Shephard, Neil, 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.

[9] Barndorff-Nielsen, Ole E., and Shephard, Neil, 2004, Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression and Correlation in Financial Economics, *Econometrica*, 72(3), 885-925.

[10] Barndorff-Nielsen, O.E., and Shephard, N., 2006, Econometrics of testing for jumps in Financial Economics using Bipower Variation, *Journal of Financial Econometrics*, 4(1), 1-30.

[11] Barndorff-Nielsen, Ole E., Nielsen, Bent, Shephard, Neil and Ysusi, Carla, 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.

[12] Barndorff-Nielsen, Ole E., Hansen, Peter R., Lunde, Asger, and Shephard, Neil, 2006, Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise, mimeo, Department of Economics, University of Oxford.

[13] Bollerslev, Tim, 1986, Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, 31, 307-327.

[14] Bollerslev, Tim, Engle, Robert F., and Nelson, Daniel B., 1994, ARCH Models, in the *Handbook of Econometrics*, R.F. Engle and D. McFadden ed.s, North Holland Press, Amsterdam.

[15] Diebold, Francis X., and Mariano, Roberto S., 1995, Comparing Predictive Accuracy, *Journal of Business and Economic Statistics*, 13(3), 253-263.

[16] Engle, Robert F., 1982, Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of U.K. Inflation, *Econometrica,* 50, 987-1008.

[17] Engle, Robert F., and Patton, Andrew J., 2001, What Good is a Volatility Model?, *Quantitative Finance*, 1(2), 237-245.

[18] Fleming, Jeffrey, Kirby, Chris, and Ostdiek, Barbara, 2001, The Economic Value of Volatility Timing, *Journal of Finance*, 56, 329-352.

[19] Foster, Dean P., and Nelson, Dan B., 1996, Continuous Record Asymptotics for Rolling Sample Variance Estimators, *Econometrica*, 64(1), 139-174.

[20] Gatheral, Jim, and Oomen, Roel, 2007, Zero-Intelligence Realized Variance Estimation, working paper, Department of Finance, Warwick Business School.

[21] Gonçalves, Sílvia, and Meddahi, Nour, 2005, Bootstrapping Realized Volatility, working paper, Université de Montréal.

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

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

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

[25] Hansen, Peter R., and Lunde, Asger, 2006a, Realized Variance and Market Microstructure Noise, *Journal of Business and Economic Statistics*, 24(2), 127-161. With comments and rejoinder.

[26] Hansen, Peter R., and Lunde, Asger, 2006b, Consistent Ranking of Volatility Models, *Journal of Econometrics*, 131, 97-121.

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

[28] Large, Jeremy, 2005, Estimating Quadratic Variation When Quoted Prices Change by a Constant Increment, working paper, Department of Economics, University of Oxford.

[29] Oomen, Roel C.A., 2006, Properties of Realized Variance under Alternative Sampling Schemes, *Journal of Business and Economic Statistics*, 24(2), 219-237.

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

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

[32] Patton, Andrew J., 2006, Volatility Forecast Comparison using Imperfect Volatility Proxies, Quantitative Finance Research Centre, University of Technology Sydney, Research Paper 175.

[33] Patton, Andrew J., and Timmermann, Allan, 2003, Properties of Optimal Forecasts under Asymmetric Loss and Nonlinearity, *Journal of Econometrics*, forthcoming.

[34] Patton, Andrew J. and Sheppard, Kevin, 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.

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

[36] Poon, Ser-Huang, and Granger, Clive W.J., 2003, Forecasting Volatility in Financial Markets, *Journal of Economic Literature*, 41(2), 478-539.

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

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

[39] Zhang, Lan, 2006, Efficient Estimation of Stochastic Volatility using Noisy Observations: A Multi-Scale Approach, *Bernoulli*, forthcoming.

[40] Zhang, Lan, Mykland, Per A., and Aït-Sahalia, Yacine, 2005, A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data, Journal of the American Statistical Association, 100(472), 1394-1411.

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

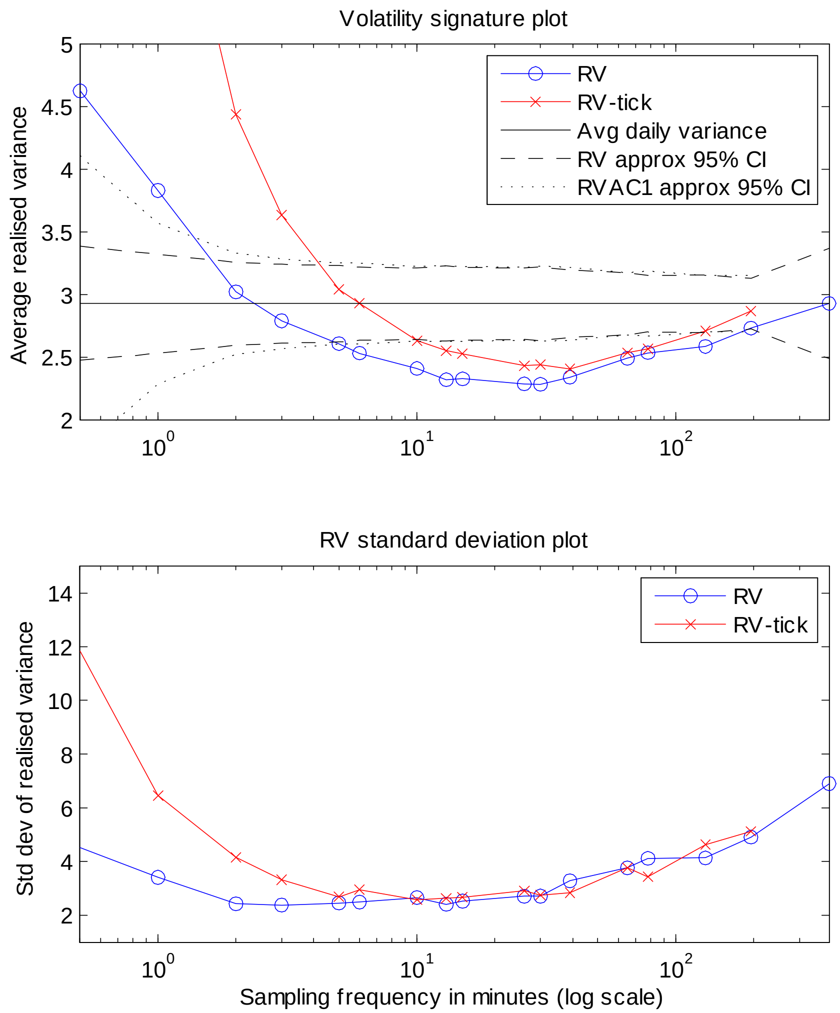

Figure 1: *Volatility signature plot (top panel) for RV and RV-tick estimators, and plot of RV and RV-tick standard deviation (lower panel). The 95% confidence interval (CI) in the upper panel is a CI for the daily squared return at the far right, and a CI for the mean difference between the daily squared return and the RV(m) or RVtick(m) estimator for the remaining values of m.*

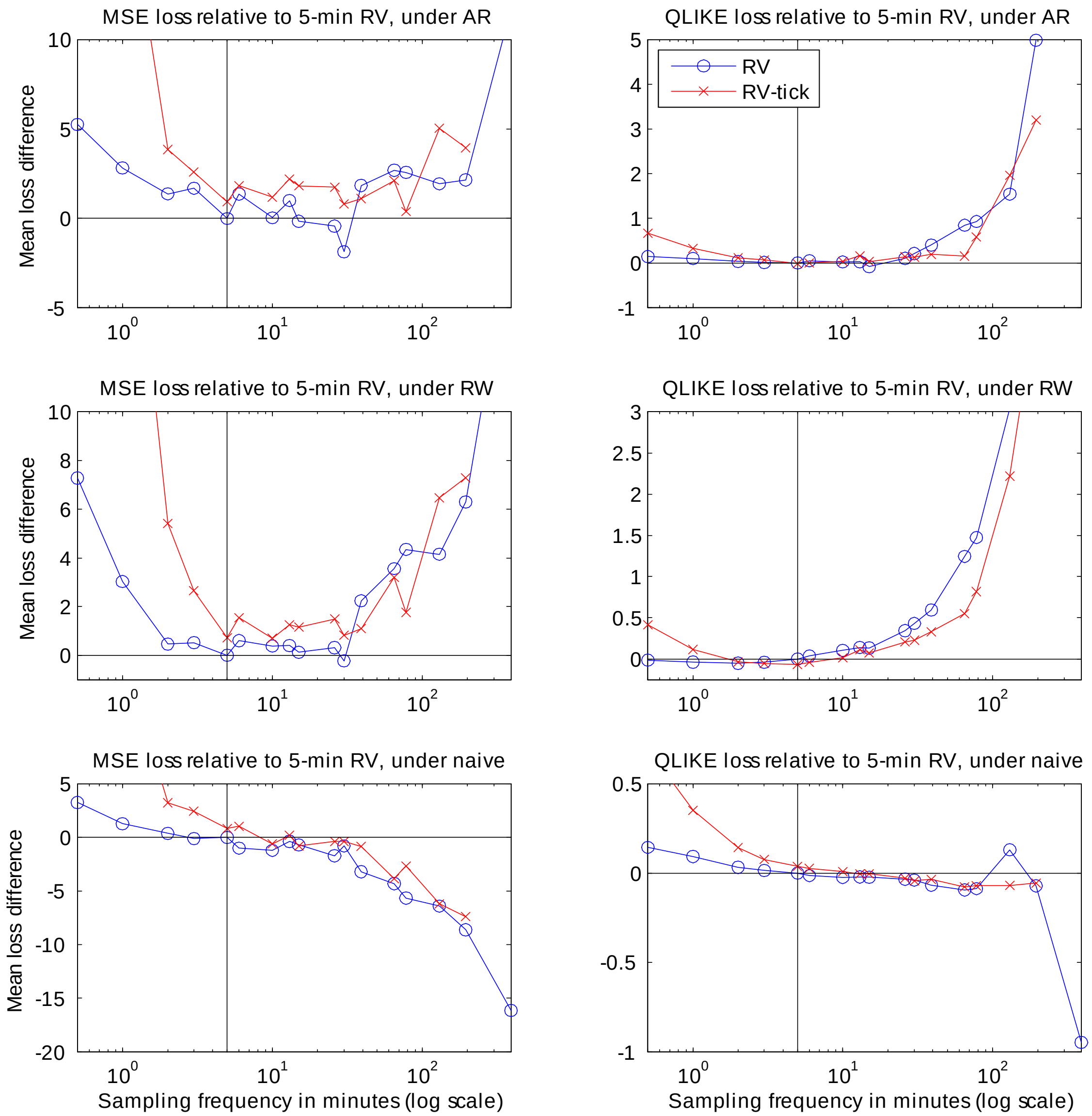

Figure 2: *Differences in average loss across sampling frequencies for the standard RV estimator and the RV-tick estimator, for MSE and QLIKE loss, according to different assumptions about the DGP. A negative loss differential means that the estimator out-performs the 5-min RV estimator.*