Data-Based Ranking of Realised Volatility Estimators

Andrew Patton

Department of Economics, and
Oxford-Man Institute of Quantitative Finance,
University of Oxford

May 2008

Outline of the talk

1. Introduction and overview of realised volatility
2. Comparisons of RV estimators in the literature
3. Data-based ranking of RV estimators
4. Application to measuring IBM equity return volatility
5. Summary and outline of future work

Background literature

- In the past 5-10 years there has been an explosion in financial econometrics research focussed on volatility *measurement* (as distinct from forecasting).
- These papers all focus on various aspects of the problem of measuring the (say) volatility of daily returns using *intra-daily* data:

Background literature, cont'd

Measuring volatility using high-frequency data

Aït-Sahalia, Mykland and Zhang (2005, RFS, 2005, JASA)

Andersen, Bollerslev, Diebold and Labys (2003, Econometrica)

Bandi and Russell (2008, REStud)

Barndorff-Nielsen and Shephard (2002, JRSS, 2004, Etca, 2004, J. F.Ects)

Hansen and Lunde (2006, JBES)

Recent surveys

Andersen, Bollerslev, Christoffersen and Diebold (2005, H'book Econ.For.)

Barndorff-Nielsen and Shephard (2007, ES monograph)

'Older' papers in this area

Andersen and Bollerslev (1998, IER)

French, Schwert and Stambaugh (1987, JFE)

Merton (1980, JFE)

A few different RV estimators

$$
\begin{aligned}
RV_t^{(m)} &= \sum_{j=1}^{m} r_{t,j}^2 \\
RVACq_t^{(m)} &= \sum_{j=1}^{m} r_{t,j}^2 + \sum_{h=1}^{q} \sum_{j=1}^{m} \left(r_{t,j} r_{t,j-h} + r_{t,j} r_{t,j+h} \right) \\
RVK_t^{(m)} &= \sum_{j=1}^{m} r_{t,j}^2 + \sum_{h=1}^{q} k \left(\frac{h-1}{q} \right) \left\{ \sum_{j=1}^{m} \left(r_{t,j} r_{t,j-h} + r_{t,j} r_{t,j+h} \right) \right\} \\
RVtick_t^{(m)} &= \sum_{k=1}^{m} r_{t,k}^2
\end{aligned}
$$

- Under various conditions, these estimators are consistent and/or unbiased for the latent quadratic variation or integrated variance:

$$
QV_t \equiv p \lim_{m \to \infty} \sum_{j=1}^{m} r_{t,j}^2 \quad , \quad IV_t \equiv \int_{t-1}^{t} \sigma^2 (s) \, ds
$$

Choosing a RV estimator: economic loss functions

- The previous contains just a few of the many RV estimators in the literature - how should one choose a particular RV for application?
- The ideal case would be to use an *economic loss function*, which describes the economic costs of estimation error in a given application:
  - **derivatives pricing**: squared pricing errors, profits from a trading strategy
  - **risk management**: costs of VaR violations, costs of holding excess capital.
  - **portfolio decisions and relative-value trading**: realised utility from portfolio, risk-adjusted returns on strategy.

Choosing a RV estimator: statistical loss functions

- In most academic studies, the economic loss function of the end-user is unknown, and so a simple statistical loss function is employed.
- The most widely-used statistical loss function is MSE:

$$L\left(IV_t, RV_t\right) = \left(IV_t - RV_t\right)^2$$

  - If the estimator is unbiased, then this measures the variance of the estimator, else it captures a bias-variance trade-off.
- Of course, we could also consider other measures of distance
- The key difficulty here, as in volatility forecasting, is that the target variable $\left(IV_t\right)$ is *unobservable*. So how do we measure accuracy?

Comparisons of RV estimators in the literature

1. **"Standard" RV theory**: choose m as large as possible
2. **Zhou (1996):** assuming *iid* noise, derived MSE-optimal choice of m for standard RV
3. **Aït-Sahalia, Mykland and Zhang (2005):** derived expressions for the MSE-optimal choice of m, for standard RV, under *iid* noise, serially correlated noise and endogenous noise
4. **Hansen and Lunde (2006)**: assuming *iid* noise, derived expression for optimal m for $RVACq$ estimators

Comparisons of RV estimators in the literature, cont'd

5. **Oomen (2006)**: assuming a parametric "pure jump" DGP, compared calendar-time returns versus "tick time" returns

6. **Andersen, Bollerslev and Meddahi (2007)**: assuming *iid* noise (possibly more), derived expression for optimal m for RV estimators, and compared $RVAC1$, RVK and the 2-scale estimator of ZMA

7. **Bandi and Russell (2006)**: assuming *iid* noise, derived expressions for the MSE-optimal choice of q/m in a $RVACq$ estimator

8. **Bandi, Russell and Yang (2007)**: derived expressions for the MSE-optimal choice of m for a standard RV estimator, assuming mean-zero but heteroskedastic noise

Motivations for a *data-based* ranking method

- In contrast with previous comparisons, the proposed methods avoid the need to take a stand on important properties of the price process. e.g., there is no need to take a stand on the particular form of noise:
    - *iid* *vs*. correlated with efficient price, see Hansen and Lunde (2006) and Kalnina and Linton (2007)
    - constant vs. time-varying noise variance, see Bandi, Russell and Yang (2007)

  ⇒ This approach *does* require assumptions on the time series properties of variables under analysis, and so this approach is a complement rather than a substitute for existing methods.

- Further, the proposed method avoids the need to estimate quantities like integrated quarticity or the variance of the noise process

- Finally, a data-based ranking method allows for comparisons that are hard/impossible using existing theory:

Motivations for a *data-based* ranking method, cont'd

Comparisons that are hard/impossible using existing theory:

- RV based on trades vs. mid-quote prices

Motivations for a *data-based* ranking method, cont'd

Comparisons that are hard/impossible using existing theory:

- RV based on trades vs. mid-quote prices

  → Theoretical comparisons would require assumptions on the quote updating process, the issuance of market vs. limit orders, etc.

Motivations for a *data-based* ranking method, cont'd

Comparisons that are hard/impossible using existing theory:

- RV based on trades vs. mid-quote prices
  - → Theoretical comparisons would require assumptions on the quote updating process, the issuance of market vs. limit orders, etc.
- RV based on calendar time vs. transaction time sampling

Motivations for a *data-based* ranking method, cont'd

Comparisons that are hard/impossible using existing theory:

- RV based on trades vs. mid-quote prices
  - $\longrightarrow$ Theoretical comparisons would require assumptions on the quote updating process, the issuance of market vs. limit orders, etc.
- RV based on calendar time vs. transaction time sampling
  - $\longrightarrow$ Theoretical comparisons require assumptions on the arrival rate of trades and/or quotes, see Oomen (2006)

Motivations for a *data-based* ranking method, cont'd

Comparisons that are hard/impossible using existing theory:

- RV based on trades vs. mid-quote prices
  - → Theoretical comparisons would require assumptions on the quote updating process, the issuance of market vs. limit orders, etc.
- RV based on calendar time vs. transaction time sampling
  - → Theoretical comparisons require assumptions on the arrival rate of trades and/or quotes, see Oomen (2006)
- The "multi-scale" RV estimator of Zhang (2006) vs. the 'alternation' estimator of Large (2005)

Motivations for a *data-based* ranking method, cont'd

Comparisons that are hard/impossible using existing theory:

- RV based on trades vs. mid-quote prices
  - → Theoretical comparisons would require assumptions on the quote updating process, the issuance of market vs. limit orders, etc.
- RV based on calendar time vs. transaction time sampling
  - → Theoretical comparisons require assumptions on the arrival rate of trades and/or quotes, see Oomen (2006)
- The "multi-scale" RV estimator of Zhang (2006) vs. the 'alternation' estimator of Large (2005)
  - → Comparisons of estimators such as these would require some way of linking their underlying assumptions

Contributions of this paper

- The primary contribution of this paper is to present a method to consistently estimate

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

- With such an estimator, many standard forecast comparison tests can then be employed:

  1. Diebold-Mariano (1995), West (1996): *pair-wise comparisons*
  2. White (2000), Hansen (2005): *comparisons of many RV estimators*
  3. Romano-Wolf (2005): *'step-wise' tests of RV estimators*
  4. Hansen-Lunde-Nason (2005): *'model confidence sets'*
  5. Giacomini-White (2006): *conditional comparisons of RV estimators*

Contributions of this paper - theory

1. I propose a formal data-based method to rank RV estimators in terms of their average distance from the latent target variable.

    1. This method employs an instrumental variables-type estimator
    2. A bias term is identified and an estimator of it is proposed
    3. I provide conditions under which existing tests in the forecast comparison literature can be used to rank RV estimators

Contributions of this paper - empirical

2. I implement these methods using high frequency data on IBM from 1996-2007, and I find:

    1. Significant gains from using prices sampled at between 15 seconds and 2 minutes, relative to daily or 5-minute prices.
    2. Tick-time sampling is preferred to calendar-time sampling, especially when trades are irregularly-spaced
    3. Transaction prices are preferred to quote prices in the early part of the sample period, but there is no difference in the latter period.

Notation

θ_t	the $\mathcal{F}_t$-meas. latent target variable, eg: QV_t or IV_t
X_{it}, $i = 1, 2, ..., n$	the $\mathcal{F}_t$-meas. realised volatility estimators
m	the number of intra-daily observations
T	the number of daily observations
$L(\theta, X)$	the pseudo-distance measure
$\tilde{\theta}_t$	a $\mathcal{F}_t$-meas., noisy, but unbiased estimator of θ_t
Y_t	the proxy or instrument for θ_t

The pseudo-distance measure

- I rank RV estimators using the average distance between the estimator and the quantity of interest:

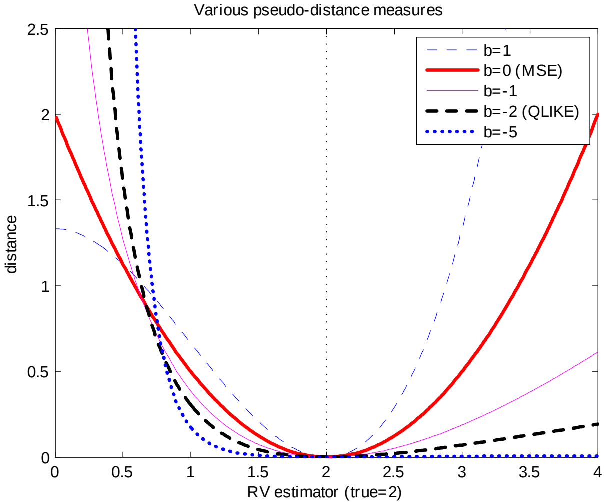
$$\begin{array}{lccc} \text{Infeasible} & E\left[L\left(\theta_t, X_{it}\right)\right] & \gtreqless & E\left[L\left(\theta_t, X_{jt}\right)\right] \\ \text{Feasible} & E\left[L\left(Y_t, X_{it}\right)\right] & \gtreqless & E\left[L\left(Y_t, X_{jt}\right)\right] \end{array}$$

  where Y_t is the proxy for θ_t.

- I use the class of pseudo-distance measures proposed in Patton (2006):

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

Distance measures



Correlated measurement errors cause problems

- From Hansen and Lunde (2006) and Patton (2006), if

$$Cov_{t-1}\left[X_t - \theta_t, \tilde{\theta}_t - \theta_t\right] = 0$$

  then MSE rankings using $\tilde{\theta}_t$ are equivalent to those using θ_t.

  - ⋆ *e.g.*: $\theta_t \equiv V_{t-1}[r_t]$, $X_t \equiv \hat{V}_{t-1}[r_t]$, and $\tilde{\theta}_t \equiv r_t^2$.

- But if

$$Cov_{t-1}\left[X_t - \theta_t, \tilde{\theta}_t - \theta_t\right] \neq 0$$

  then MSE rankings using $\tilde{\theta}_t$ are **not** equivalent to those using θ_t.

  - ⋆ The fact that $(\theta_t, X_t) \notin \mathcal{F}_{t-1}$ in RV comparison causes problems..

- I will break this correlation in a familiar way:

IV estimation for IV comparison

- I will overcome the problem of correlated measurement errors:

$$Cov_{t-1}\left[X_t - \theta_t, \tilde{\theta}_t - \theta_t\right] \neq 0$$

in a standard way, by using a *lead* of the proxy:

$$Y_t = \tilde{\theta}_{t+1} = \theta_{t+1} + \nu_{t+1}.$$

This approach exploits two features of the problem:

1. The target variable (IV or QV) is known to be persistent, so θ_{t+1} is highly correlated with θ_t
2. Almost all RV estimators in the literature are one-sided in nature: X_t uses data only up until day t (and usually *only* data from day t). So measurement error in X_t is uncorrelated with meas error in $\tilde{\theta}_{t+1}$

IV estimation for IV comparison, cont'd

- This problem is a *non-linear* instrumental variables problem, and we need to put more structure on the problem than just non-zero correlation.
  - It is not sufficient to simply assume $Cov\left[\tilde{\theta}_{t+1} - \theta_{t+1}, X_t - \theta_t\right] = 0$ and $Cov\left[\tilde{\theta}_{t+1}, \theta_t\right] \neq 0$
- I will consider approximating the conditional mean of the target variable using two approaches:
  1. A random walk approximation
  2. A general (stationary) AR(p) approximation
- I will show via simulation that both these models are reasonable approximations for a realistic DGP.

A random walk approximation for the target variable

- Numerous papers on the conditional variance or integrated variance have reported that these quantities are very persistent, close to being random walks.
  - Eg: The widely-used RiskMetrics model is based on a unit root assumption for the conditional variance.
  - See Bollerslev, *et al.* (1994), Andersen, *et al.*, (2003, 2005), amongst many others, on the behaviour of conditional volatility
  - Note that Wright (1999) provides evidence *against* the presence of a unit root in daily conditional variance for many stocks.
- Given this, consider the following assumption:

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

Assumptions for the RW approximation

- The standard conditional unbiasedness assumption for the noisy proxy:

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

- It is simple to consider convex combinations of *leads* of $\tilde{\theta}_t$ as our proxy:

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 $\sum_{i=1}^{J} \omega_i = 1$.

Rankings based on a RW approximation

Proposition

(a) Let assumptions T1, P1 and P2 hold. Then:

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

for any vector of RV estimators, $\mathbf{X}_t$.

Rankings based on a RW approximation

- The intuition behind this result is based on:

$$\begin{aligned} \tilde{\theta}_{t+1} &= \theta_{t+1} + \nu_{t+1} \\ &= \theta_t + \eta_{t+1} + \nu_{t+1} \\ &\equiv \theta_t + \epsilon_{t+1} \\ \text{with} \quad Corr\left[\epsilon_{t+1}, X_t\right] &= 0 \end{aligned}$$

- Thus if θ_t is very persistent, then tomorrow's *proxy*, $\tilde{\theta}_{t+1}$ is a good estimate of today's target variable θ_t.

- Next, I draw on existing work on forecast comparison to obtain a distribution theory for the feasible estimate of the differences in distances.

Rankings based on a RW approximation, cont'd

Proposition

(b) If we further assume mixing and moment conditions (A1 and A2), then:

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

Rankings based on a RW approximation, cont'd

Proposition

(c) If $p_T \to 0$ and $T \times p_T \to \infty$ as $T \to \infty$, where p_T is the inverse of the average block length in Politis and Romano's (1994) stationary bootstrap, then the stationary bootstrap may also be employed, as:

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

An AR(p) approximation for the target variable

- Meddahi (2003, *EJ*) and Barndorff-Nielsen (2002, *JRSS-B*) show that the integrated variance follows an ARMA(p,q) model for a wide variety of stochastic volatility models for the spot volatility.
  - Eg: Meddahi shows that a p-factor SV model generates an ARMA(p, p) for the daily integrated variance
- Empirical and theoretical work by Andersen, Bollerslev and Meddahi (2004 *IER*, 2007 *wp*) reveals that an AR(1) performs no worse than the optimal ARMA(p,q) model for a range of realistic DGPs.
- The result below may be generalised to hold for invertible ARMA(p,q) processes, but in light of the empirical work in this area, I consider only AR(p) processes.

Rankings based on an AR(p) approximation

- The following assumption allows the target variable to follow (almost) any stationary AR(p) process:

Assumption T2:

$$\begin{aligned} \theta_t &= \phi_0 + \sum_{i=1}^{p} \phi_i \theta_{t-i} + \eta_t, \\ E\left[\eta_t | \mathcal{F}_{t-1}\right] &= 0 \end{aligned}$$

with $\phi_1 \neq 0$ and $\Phi \equiv \left[\phi_0, \phi_1, ..., \phi_p\right]'$ such that θ_t is covariance stationary.

- The following result uses an instrumental variables estimator to obtain the AR(p) parameters for θ_t.

Rankings based on an AR(p) approximation

Proposition

(a) Let assumptions T2, P1 and P2 hold, and let R2 hold if $p > 1$. Then

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

$$\text{where } \beta = \frac{\phi_0}{\phi_1} E\left[\Delta C\left(\mathbf{X}_t; b\right)\right] + \left(1 - \frac{1}{\phi_1}\right) E\left[\Delta C\left(\mathbf{X}_t; b\right) Y_t\right] + \sum_{i=2}^{p} \frac{\phi_i}{\phi_1} E\left[\Delta C\left(\mathbf{X}_t; b\right) Y_{t-i}\right]$$

Rankings based on an AR(p) approximation, cont'd

Proposition

(b) If we further assume mixing and moment conditions (A1 and A2), then:

$$\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}\Delta L\left(Y_t,\mathbf{X}_t;b\right)-\hat{\beta}_T-E\left[\Delta L\left(\theta_t,\mathbf{X}_t;b\right)\right]\right)\to^{d} N\left(0,\Omega\right)$$

Rankings based on an AR(p) approximation, cont'd

Proposition

(c) If $p_T \to 0$ and $T \times p_T \to \infty$ as $T \to \infty$ then the stationary bootstrap may also be employed, as:

$$\sup_z \left| P^* \left[\left\| \frac{1}{T} \sum_{t=1}^{T} \Delta L \left(Y_t^*, \mathbf{X}_t^*; b \right) - \hat{\beta}_T^* - \frac{1}{T} \sum_{t=1}^{T} \Delta L \left(Y_t, \mathbf{X}_t; b \right) + \hat{\beta}_T \right\| \leq z \right] \right.$$
$$\left. - P \left[\left\| \frac{1}{T} \sum_{t=1}^{T} \Delta L \left(Y_t, \mathbf{X}_t; b \right) - \hat{\beta}_T - E \left[\Delta L \left(\theta_t, \mathbf{X}_t; b \right) \right] \right\| \leq z \right] \right| \to 0$$

Conditional rankings of RV estimators

- The final theoretical result in the paper is to consider *conditional* comparisons of RV estimators, using the framework of Giacomini and White (2006).
- The null hypothesis in a GW-type test is:

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

- The above null is usually tested by looking at simple regressions of the form:

$$\Delta L\left(\theta_t, \mathbf{X}_t\right) = \boldsymbol{\alpha}^{*\prime}\mathbf{Z}_{t-1} + e_t^*$$

where $\mathbf{Z}_{t-1} \in \mathcal{G}_{t-1}$ is some vector of variables, and then testing:

$$\begin{aligned} H_0' &: \boldsymbol{\alpha}^* = 0 \\ \text{vs.} \quad H_a' &: \boldsymbol{\alpha}^* \neq 0 \end{aligned}$$

Conditional rankings of RV estimators, cont'd

- Infeasible regression:

$$\Delta L\left(\theta_t, \mathbf{X}_t\right) = \boldsymbol{\alpha}^{*\prime}\mathbf{Z}_{t-1} + e_t^*$$

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

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

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

Conditional rankings of RV estimators

Proposition

(a) Let assumptions T1, P1 and P2 hold. Then

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

for any vector of RV estimators, $\mathbf{X}_t$.

Conditional rankings of RV estimators, cont'd

Proposition

(b) Denote the OLS estimator of $\boldsymbol{\alpha}$ *as* $\hat{\boldsymbol{\alpha}}_T$*. Then under mixing and moment conditions (A3 and A4):*

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

$$\begin{aligned} \textit{where} \quad \hat{D}_T &\equiv \hat{M}_T^{-1}\hat{\Omega}_T\hat{M}_T^{-1} \\ \hat{M}_T &\equiv \frac{1}{T-1}\sum_{t=2}^{T}\mathbf{Z}_{t-1}\mathbf{Z}_{t-1}' \\ \Omega_T &\equiv V\left[\frac{1}{\sqrt{T-1}}\sum_{t=2}^{T}\mathbf{Z}_{t-1}e_t\right] \end{aligned}$$

and with $\hat{\Omega}_T$ *some estimator such that* $\hat{\Omega}_T - \Omega_T \rightarrow^p 0$.

A small simulation study - the DGP

- To check the finite-sample size properties of the proposed methods, I conducted a small simulation study:
- I use a standard log-normal stochastic volatility model with a leverage effect, with the same parameters as in Goncalves and Meddahi (2005):

$$d \log P_t^* = 0.0314 dt + \nu_t \left(-0.576 dW_{1t} + \sqrt{1 - 0.576^2} dW_{2t} \right)$$

$$d \log \nu_t^2 = -0.0136 \left(0.8382 + \log \nu_t^2 \right) dt + 0.1148 dW_{1t}$$

- 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).
- I look at sequences of 500 and 2500 'trade days'.

Simulation design - adding some noise

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

$$\begin{aligned} \log P\left(t_j\right) &= \log P^*\left(t_j\right) + \xi\left(t_j\right) \\ \xi\left(t_j\right) &\sim iid\;\; N\left(0, \sigma_\xi^2\right) \\ \text{where } \frac{2\sigma_\xi^2}{V\left[r_t\right]\frac{5}{390} + 2\sigma_\xi^2} &= 0.20 \end{aligned}$$

- i.e., the variance of the noise is 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)
  - The proportion of 20% is around the middle value considered in the simulation study of Huang and Tauchen (2005).

Goodness-of-fit of ARMA models for IV

- Meddahi (2003) and Barndorff-Nielsen and Shephard (2002) show theoretically that integrated variance follows an ARMA(p,q) model for a wide variety of stochastic volatility models for the instantaneous volatility (though they assume no noise and no leverage effect)

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

Simulation design - the competing RV estimators

- Next I consider the finite-sample size of pair-wise comparisons obtained via a bootstrap verison of a Diebold-Mariano (1995) test.

- I set the each RV estimator equal to the true IV plus some noise:

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

Simulation design - the competing RV estimators

- Next I consider the finite-sample size of pair-wise comparisons obtained via a bootstrap verison of a Diebold-Mariano (1995) test.

- I set the each RV estimator equal to the true IV plus some noise:

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

- I set $Corr\left[\nu_t^{30\,\text{min}}, \zeta_{1t}\right] = 0.5$.

Simulation design - the competing RV estimators

- Next I consider the finite-sample size of pair-wise comparisons obtained via a bootstrap verison of a Diebold-Mariano (1995) test.

- I set the each RV estimator equal to the true IV plus some noise:

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

- I set $Corr\left[\nu_t^{30\,\text{min}}, \zeta_{1t}\right] = 0.5$.

- In the study of the size of the tests I set $\sigma_{\zeta 1}^2 = \sigma_{\zeta 2}^2 = 0.1 \times V[IV_t]$. To study the power, I fix $\sigma_{\zeta 1}^2$,and let $\sigma_{\zeta 2}^2 / V[IV_t] = 0.15,\ 0.2,\ 0.5,\ 1$.

Finite-sample size and power, T=500, using MSE

		IV		*RV-30min*		*RV-daily*	
γ	*IV**	R.W.	AR(1)	R.W.	AR(1)	R.W.	AR(1)
0.10	0.05	0.03	0.02	0.04	0.00	0.06	0.01
0.15	0.98	0.89	0.88	0.40	0.02	0.14	0.00
0.20	1.00	1.00	1.00	0.74	0.06	0.23	0.02
0.50	1.00	1.00	1.000	1.00	0.56	0.60	0.06
1.00	1.00	1.00	1.000	1.00	0.70	0.89	0.07

Simulation design - conditional comparisons

- Next I consider a simple design to check the finite-sample size of GW tests for this application.

$$\begin{aligned} X_{1t} &= IV_t + \zeta_{1t} \\ X_{2t} &= IV_t - \lambda IV_{t-1} + \zeta_{2t} \\ \zeta_{it} &= \omega v_t^{30\,\text{min}} + (1-\omega)\,\sigma_u U_{it}, \quad i = 1, 2 \\ [U_{1t}, U_{2t}]' &\sim iid\ N(0, I) \end{aligned}$$

Simulation design - conditional comparisons

- Next I consider a simple design to check the finite-sample size of GW tests for this application.

$$\begin{aligned} X_{1t} &= IV_t + \zeta_{1t} \\ X_{2t} &= IV_t - \lambda IV_{t-1} + \zeta_{2t} \\ \zeta_{it} &= \omega v_t^{30\,\text{min}} + (1-\omega)\,\sigma_u U_{it}, \quad i = 1,2 \\ [U_{1t}, U_{2t}]' &\sim iid\;\; N(0, I) \end{aligned}$$

- To study finite-sample size, I set $\lambda = 0$. To study power, set $\lambda = 0.1, 0.2, 0.4, 0.8$. Tests are based on regressions of the form:

$$\begin{aligned} L\left(\tilde{\theta}_{t+1}, X_{1t}\right) - L\left(\tilde{\theta}_{t+1}, X_{2t}\right) &= \alpha_0^u + e_t^u, \quad \text{or} \\ L\left(\tilde{\theta}_{t+1}, X_{1t}\right) - L\left(\tilde{\theta}_{t+1}, X_{2t}\right) &= \alpha_0 + \alpha_1 \log \frac{1}{10} \sum_{j=1}^{10} \tilde{\theta}_{t-j} + e_t \end{aligned}$$

Finite-sample size and power, T=500, using MSE

	Conditional - slope test			*Conditional - joint test*		
		Volatility proxy			Volatility proxy	
λ	IV*	IV	RV-daily	IV*	IV	RV-daily
0	0.06	0.08	0.04	0.06	0.06	0.04
0.1	0.28	0.16	0.05	0.33	0.21	0.05
0.2	0.93	0.80	0.05	0.92	0.86	0.08
0.4	1.00	1.00	0.11	1.00	1.00	0.34
0.8	1.00	1.00	0.47	1.00	1.00	0.88

Summary of simulation results

- For a realistic DGP, with noise and a leverage effect, I find that the finite-sample size is reasonable, with rejection frequencies close to 0.05.
- The results for the power of the tests are as expected:
  1. power of the new tests are worse than would be obtained if IV were observable
  2. power is worse when a noisier instrument is used (daily squared returns versus 30-minute RV versus true IV)
  3. power of the tests based on the AR(1) assumption are worse than those based on the random walk assumption.
  4. power of the tests are better when a larger sample size is available

Application to IBM stock returns

- I consider estimating the quadratic variation of the daily return on IBM, using data from TAQ from Jan 1996 to June 2007, yielding 2893 daily observations.
    - I break this sample into three sub-periods (1996-1999, 2000-2003, 2004-2007) to allow for changes in market rules and conditions.
- I use standard RV, based on:
    1. trade prices and mid-quote prices
    2. calendar-time sampling and tick-time sampling
    3. sampling frequencies of 1, 2, 5, 15, 30 seconds, 1, 2, 5, 15, 30 minutes, 1, 2 hours and 1 day.
- The total number of RV estimators is $2 \times 2 \times 13 - 4 = 48$

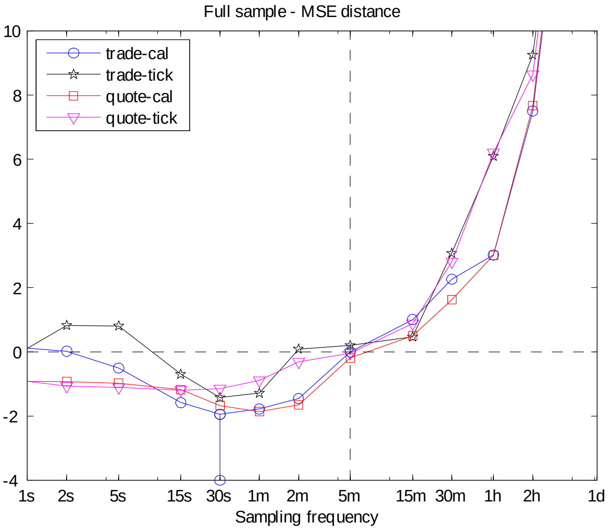
Data-based comparisons of the 48 RV estimators

1. Raw rankings of the RV estimators based on estimated average differences in distance

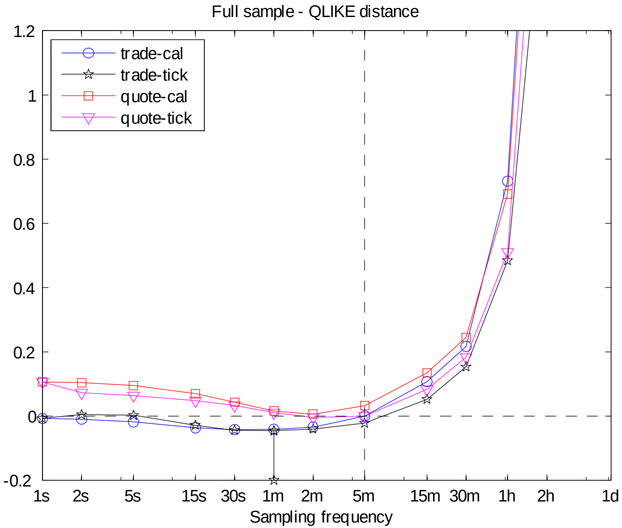
2. The stepwise multiple testing method of Romano-Wolf (2005)
   - Which estimators significant beat (or are beaten by) **daily RV**?
   - Which estimators significant beat (or are beaten by) **5-minute RV**?

3. The conditional comparison test of Giacomini-White (2006)
   - Does high frequency data help more during **volatile periods**?
   - When are quote prices more or less informative than transaction prices?
   - Does tick-time sampling help when **trades arrive irregularly**?

Estimated differences in distance under MSE



Estimated differences in distance under QLIKE



The Romano-Wolf stepwise test

- The Romano-Wolf test looks at each of 47 null and alternative hypotheses separately:

$$
\begin{aligned}
H_0^{(i)} &: \quad E\left[L\left(\theta_t, X_{0t}\right) - L\left(\theta_t, X_{it}\right)\right] \leq 0 \\
H_1^{(i)} &: \quad E\left[L\left(\theta_t, X_{0t}\right) - L\left(\theta_t, X_{it}\right)\right] > 0
\end{aligned}
$$

  and identifies *which* null hypotheses can be rejected.

- Romano-Wolf's procedure controls the 'family-wise error rate' of these 47 tests
  - FWE is the probability that we reject at least one true null hypothesis, and reduces to the size of the test if we examine only one null.

The Romano-Wolf stepwise test - results

Daily RV on transaction prices as the benchmark

MSE			QLIKE		
Better	Not Diff	Worse	Better	Not Diff	Worse
33	14	0	46	1	0

- Under QLIKE, daily RV is significantly beaten by every other estimator, except for daily RV using quote prices.
- Under MSE it is beaten by 33 estimators. Those that do *not* beat it are *RV* using 30-min or lower sampling.

The Romano-Wolf stepwise test - results

5-minute calendar-time RV on transaction prices as the benchmark

MSE			QLIKE		
Better	Not Diff	Worse	Better	Not Diff	Worse
0	47	0	9	9	29

- Under MSE, no estimator can be distinguished from 5-min RV. (Power problem with this application.)
- Under QLIKE, most estimators are *worse* than 5-min RV, but a few are significantly better: those based on trade prices sampled at between 15 seconds and 5 minutes.

High-frequency vs. Low-frequency RV estimators

Conditional on recent volatility

$$L\left(Y_t, RV_t^{daily}\right) - L\left(Y_t, RV_t^{5\,\text{min}}\right) = \underset{(8.75)}{36.14} + e_t$$

$$L\left(Y_t, RV_t^{daily}\right) - L\left(Y_t, RV_t^{5\,\text{min}}\right) = \underset{(10.70)}{26.71} + \underset{(2.94)}{19.20} Z_{t-1} + e_t$$

$$\text{where} \quad Z_{t-1} = \log \frac{1}{10} \sum_{j=1}^{10} \tilde{\theta}_{t-j}$$

- The positive constant in the 1st regression reveals that daily squared returns are worse than 5-min RV
- The positive and significant slope coefficient in the 2nd regression reveals that daily squared returns are particularly bad proxies during high liquidity periods. (pval on joint test is <0.000)

Tick-time vs. Calendar-time sampling

Conditional on the volatility of trade durations

$$
\begin{aligned}
L\left(Y_t, RVtick_t^{(h\min)}\right) - L\left(Y_t, RV_t^{(h\min)}\right) &= \alpha^u + e_t^u \\
L\left(Y_t, RVtick_t^{(h\min)}\right) - L\left(Y_t, RV_t^{(h\min)}\right) &= \alpha_0 + \alpha_1 Z_{t-1} + e_t \\
\text{where} \quad Z_{t-1} &\equiv V\left[\text{Duration}_{j,t-1}\right]^{1/2}
\end{aligned}
$$

- I run these regressions for each value of h :

Tick-time vs. Calendar-time sampling

Conditional on the volatility of trade durations

Frequency	Average (t-stat)	Intercept (t-stat)	Slope (t-stat)	Joint *p*-val
2 sec	0.01 * (10.81)	0.08* (3.91)	−0.01* (−3.26)	0.00
15 sec	0.00 (1.91)	−0.07 (−1.52)	0.01 (1.56)	0.14
30 sec	−0.01* (−2.90)	0.06 (1.15)	−0.01 (−1.22)	0.01
2 min	−0.01* (−3.55)	0.06 (1.65)	−0.01 (−1.80)	0.00
15 min	−0.06* (−7.94)	0.30* (1.96)	−0.06* (−2.32)	0.00
30 min	−0.08* (−4.76)	0.82* (2.37)	−0.16* (−2.57)	0.00
2 hr	−1.23* (−2.39)	17.49* (2.33)	−3.37* (−2.40)	0.03
Joint	**−0.06** (**−7.94**)	**0.07*** (**2.50**)	**−0.01*** (**−2.22**)	**0.00**

Quote prices vs. Trade prices

Conditional on the ratio of number of quotes to number of trades

$$
\begin{aligned}
L\left(Y_t, RV_t^{quote(h\min)}\right) - L\left(Y_t, RV_t^{trade(h\min)}\right) &= \alpha^u + e_t^u \\
L\left(Y_t, RV_t^{quote(h\min)}\right) - L\left(Y_t, RV_t^{trade(h\min)}\right) &= \alpha_0 + \alpha_1 Z_{t-1} + e_t \\
\text{where} \quad Z_{t-1} &\equiv \frac{\#\left\{quotes\right\}_{t-1}}{\#\left\{trades\right\}_{t-1}}
\end{aligned}
$$

- I again run these regressions for each value of h :

Quote prices vs. Trade prices

Conditional on the ratio of number of quotes to number of trades

Frequency	Average (t-stat)	Intercept (t-stat)	Slope (t-stat)	Joint *p*-val
2 sec	0.14 * (9.63)	0.38 * (10.42)	−0.16* (−8.53)	0.00
15 sec	0.13 * (11.78)	0.35 * (12.50)	−0.14 * (−10.39)	0.00
30 sec	0.10 * (12.03)	0.28 * (12.34)	−0.11 * (−10.25)	0.00
2 min	0.05 * (10.65)	0.14 * (10.00)	−0.06* (−8.47)	0.00
15 min	0.03* (6.78)	0.09* (5.88)	−0.03* (−4.58)	0.00
30 min	0.03* (4.95)	0.08* (3.67)	−0.03* (−2.70)	0.00
2 hr	−0.29 (−1.10)	−1.47 (−1.70)	0.75 (1.83)	0.17
Joint	**0.06** * (**11.54**)	**0.17*** (**11.32**)	**−0.07*** (**−9.51**)	**0.00**

Conclusion and summary of results

- This paper presents conditions under which the relative average accuracy of competing RV estimators can be consistently ($T \to \infty$) estimated from available data
  - Based on plausible assumptions about the time series properties of the data
  - No need for precise assumptions about the underlying price process or market microstructure noise process
- This "data-based" ranking approach facilitates the use of standard forecast comparison tests for ranking RV estimators:
  - Diebold-Mariano (1995), West (1996), White (2000), Hansen (2005), Romano-Wolf (2005), Hansen *et al.* (2005), Giacomini-White (2006), for example.