Optimal Combinations of Realised Volatility Estimators

Andrew Patton and Kevin Sheppard

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

March 2009

Introduction

- The development of new estimators of asset price variability has been a very active area of research in the past decade
  - See Andersen, *et al.* (2006) or Barndorff-Nielsen and Shephard (2007) for recent reviews of the literature on realised volatility estimators.
- Issues considered by papers in this area:
  - Accuracy of estimators based on higher frequency data
  - Efficiency
  - Robustness to microstructure effects
  - Ability to distinguish the continuous and the ‘jump’ components of variation
  - Estimation of covariances and correlations

Partial list of papers in this area

- French, *et al.* (1987)
- Zhou (1996)
- Andersen and Bollerslev (1998)
- Andersen, Bollerslev, Diebold and Labys (2001, 2003)
- Barndorff-Nielsen and Shephard (2004, 2004, 2006)
- Aït-Sahalia, Mykland and Zhang (2005), ZMA (2005)
- Hansen and Lunde (2006)
- Christensen and Podolskij (2007), Martens and van Dijk (2007)
- Bandi and Russell (2006, 2008)
- Christensen, Oomen and Podolskij (2008)
- Barndorff-Nielsen, Hansen, Lunde and Shephard (2009)

amongst many others

This paper's main question

★ *Do combinations of RV estimators offer gains in accuracy relative to individual estimators?*

- This question is motivated by the success of combinations in forecasting applications, see Bates and Granger (1969), Stock and Watson (2004), and Timmermann (2006) for example.
- Why do combination forecasts work well? From Timmermann (2006):
    1. Combine information from each individual forecast: *directly applies to volatility estimation*
    2. Average across differences in impact of structural breaks: *directly applies to volatility estimation*
    3. Less sensitive to model mis-specification: *applies to volatility estimation in terms of assumptions used to obtain specific estimators*

Contribution of this paper

1. We propose methods for constructing theoretically optimal combinations of RV estimators, in terms of average accuracy.
    - This problem is non-standard, as the target variable (the quadratic variation of the process) is unobservable
    - Uses an extension of the data-based ranking method in Patton (2008), which avoids the need to make strong assumptions about the underlying price process.
2. We apply these methods to a collection of 32 different realised measures, across 8 distinct classes of estimators, using data on IBM from 1996-2008.
    - We use the step-wise testing method of Romano and Wolf (2005) and the MCS of Hansen, Lunde and Nason (2005) to identify best individual estimators, and to compare them with combination estimators.
    - We compare these estimators both in terms of in-sample accuracy, and in an out-of-sample forecasting experiment.

Notation

θ_t	the $\mathcal{F}_t$-meas. latent target variable, eg: QV_t or IV_t
X_{it}, $i = 1, 2, ..., n$	the $\tilde{\mathcal{F}}_t$-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 $\tilde{\mathcal{F}}_t$-meas., noisy, but unbiased estimator of θ_t
Y_t	the proxy or instrument for θ_t

The pseudo-distance measure

- We measure accuracy 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.

- General results are given for the class of pseudo-distance measures proposed in Patton (2006). Empirical results use MSE or QLIKE:

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

Combinations of RV estimators

- Let $\mathbf{X}_t = [X_{1t}, ..., X_{nt}]'$ be the vector of all n individual RV estimators, and consider a parametric combination of these:

$$
\begin{aligned}
X_t^{combo} &= g(\mathbf{X}_t; \mathbf{w}) \\
\text{eg 1} \quad g(\mathbf{X}_t; \mathbf{w}) &= w_0 + \sum_{i=1}^{n} w_i X_{it} \\
\text{eg 2} \quad g(\mathbf{X}_t; \mathbf{w}) &= w_0 \times \prod_{i=1}^{n} X_{it}^{w_i}
\end{aligned}
$$

- *Optimal* combinations:

$$
\begin{aligned}
\mathbf{w}^* &\equiv \arg\min_{\mathbf{w} \in \mathcal{W}} E\left[L\left(\theta_t, g\left(\mathbf{X}_t; \mathbf{w}\right)\right)\right] \\
\tilde{\mathbf{w}}^* &\equiv \arg\min_{\mathbf{w} \in \mathcal{W}} E\left[L\left(Y_t, g\left(\mathbf{X}_t; \mathbf{w}\right)\right)\right] \\
\hat{\mathbf{w}}_T^* &\equiv \arg\min_{\mathbf{w} \in \mathcal{W}} \frac{1}{T} \sum_{t=1}^{T} L\left(Y_t, g\left(\mathbf{X}_t; \mathbf{w}\right)\right)
\end{aligned}
$$

Assumptions for the main theoretical result

- In addition to standard regularity conditions, we require:

Assumption P1: $E\left[\tilde{\theta}_t|\theta_t, \mathcal{F}_{t-1}\right] = \theta_t$, where $\mathcal{F}_{t-1}$ is info set generated by complete path of log-price process.

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

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

- The first assumption is reasonable, if we believe squared daily returns to be noisy but unbiased estimators of QV
  - In presence of jumps some care is required to find a proxy for IV.
- The third assumption is stronger. Critical for the result is that θ_t is persistent. This can be captured either through a RW approximation (as above) or an AR approximation.

Optimal combinations of RV estimators

Proposition: Under assumptions P1, P2, T1 and regularity conditions, and if L is a member of the class of distance measures in Patton (2006), then $\tilde{\mathbf{w}}^* = \mathbf{w}^*$ and:

$$\hat{V}_T^{-1/2}\sqrt{T}\left(\hat{\mathbf{w}}_T^* - \mathbf{w}^*\right) \xrightarrow{D} N\left(0, I\right), \text{ as } T \rightarrow \infty$$

- Thus we can estimate the optimal combination parameter and overcome the fact that the target variable is latent.

Application to estimating IBM price variability

- We apply this method to estimating the variability of open-to-close returns on IBM, over Jan 1996 to July 2008, 3168 trading days.
- We consider a total of 32 individual RV estimators from 8 distinct classes of estimators
  - For each estimator we follow the implementation of the authors of the original paper as closely as possible (and in most cases exactly)
- We compare these individual estimators with 3 simple combination estimators (arithmetic mean, median, and geometric mean)
- We compare accuracy both using the previous method for estimating (in-sample) accuracy, and using an out-of-sample forecasting experiment

Description of the estimators I

1. **Realised variance**: $RV_t^{(m)} = \sum_{j=1}^{m} r_{t,j}^2$
   - Sampling frequency: 1sec, 5sec, 1min, 5min, 1hr and 1day
   - Sampling method: calendar time and tick time (Hansen and Lunde 2006, Oomen 2006)
   - Bandi and Russell's (2006, 2008) MSE-optimal frequency (in calendar time), with and without their bias correction
2. **First-order autocorrelation adjusted RV**, as in French, *et al.* (1987), Zhou (1996), Hansen and Lunde (2006), Bandi and Russell (2008)
   - Estimated on 1min and 5min returns, in calendar time.
3. **Two-scale** RV of Zhang, *et al.* (2006) and **Multi-scale** RV of Zhang (2006)
   - 1tick and 1min tick-time frequencies

Description of the estimators II

4. **Realised kernels** of Barndorff-Nielsen, *et al.* (2008), using their optimal bandwidth for each kernel
    - Kernels: Bartlett, Cubic, modified Tukey-Hanning$_2$, non-flat-top Parzen
    - 1tick and 1min tick-time sampling
5. **Realised range**-based RV of Christensen and Podolskij (2007) and Martens and van Dijk (2007)
    - Using 5min blocks, and 1min prices within each block, similar to Christensen and Podolskij.
6. **Bi-power variation** of Barndorff-Nielsen and Shephard (2006).
    - 1min and 5min sampling, in calendar time
7. **Quantile-based realised variance** of Christensen, *et al.* (2008)
    - Using quantiles of 0.85, 0.90, and 0.96. Number of sub-intervals=1.
    - Prices sampled every 1min in tick time
8. **MinRV** and **MedRV** of Andersen, *et al.* (2008):
    - Using 1min tick time sampling

Summary statistics on a sub-set of the estimators

	Mean	Standard Deviation	Skewness	Kurtosis	Minimum
$RV^{1\,\text{sec}}$	3.158	3.005	2.940	22.270	0.168
$RV^{1\,\text{min}}$	2.438	2.387	3.647	34.193	0.116
$RV^{1\text{day}}$	2.403	6.228	10.638	193.816	0.000
$RV^{AC1,1\,\text{min}}$	2.440	2.392	3.592	32.743	0.117
$TSRV^{tick}$	2.177	2.202	3.994	39.384	0.081
$MSRV^{tick}$	2.181	2.287	5.572	85.553	0.081
RK^{TH2}	2.381	2.784	7.761	158.827	0.109
RRV	2.310	2.537	4.647	52.061	0.123
$BPV^{1\,\text{min}}$	2.105	2.075	2.632	12.867	0.077
QRV	2.441	2.273	2.430	11.563	0.104
MedRV	2.260	2.157	2.600	13.216	0.109

Correlation between a sub-set of the estimators

	$RV^{1\,sec}$	$RV^{5\,min}$	RV^{1day}	$RV^{AC1\,min}$	$RK^{TH2,1\,min}$	QRV
$RV^{1\,sec}$	1	0.855	0.431	0.939	0.839	0.906
$RV^{1\,min}$	0.938	0.948	0.517	0.997	0.939	0.956
RV^{1day}	0.431	0.570	1	0.514	0.593	0.483
$RV^{AC1\,min}$	0.939	0.947	0.514	1	0.939	0.955
$TSRV^{tick}$	0.913	0.938	0.507	0.983	0.931	0.935
$MSRV^{tick}$	0.904	0.952	0.519	0.980	0.945	0.913
RK^{TH2}	0.874	0.975	0.550	0.967	0.974	0.885
RRV	0.902	0.984	0.550	0.982	0.974	0.933
$BPV^{1\,min}$	0.912	0.878	0.478	0.960	0.871	0.974
QRV	0.906	0.872	0.483	0.955	0.867	1
MedRV	0.919	0.882	0.487	0.961	0.874	0.980

In-sample performance of the estimators

The data-based ranking method of Patton (2008) requires some choices:

1. We use a one-period lead of $RV^{5\,min}$ as our instrument for the latent quadratic variation
   - Using a lower frequency (eg RV^{1day}) reduces the power of tests
   - Using a higher frequency risks violating the unbiasedness assumption
2. We will present results using QLIKE; results using MSE are in a web appendix.
   - Results are broadly similar, though power is lower using MSE than using QLIKE
3. We use the RW approximation rather than an AR approximation to the dynamics in QV
   - This was found to be satisfactory in Patton (2008) on the same data
   - AR approximation leads to similar results, though with less precision

In-sample performance of a sub-set of the estimators

	Avg ΔQLIKE	Rank				In MCS?
	Full	Full	96-99	00-03	04-08	Full
$RV^{1\,\text{sec}}$	-0.013	14	6	28	21	–
$RV^{1\,\text{min}}$	-0.040	2	3	3	1	✓
$RV^{1\text{day}}$	29.191	35	35	35	35	–
$RV^{AC1,1\,\text{min}}$	-0.040	1	2	2	2	✓
$TSRV^{tick}$	-0.001	22	27	15	20	–
$MSRV^{tick}$	-0.003	21	26	17	19	–
RK^{TH2}	-0.014	12	16	11	9	–
RRV	-0.016	9	15	8	16	–
$BPV^{1\,\text{min}}$	0.029	28	31	12	8	–
QRV	-0.035	4	4	5	4	–
MedRV	-0.024	6	9	6	10	–
RV^{Mean}	-0.030	5	8	4	3	–
$RV^{Geo-mean}$	-0.015	10	12	13	18	–
RV^{Median}	-0.020	8	11	7	6	–

In-sample Romano-Wolf tests

Benchmark: Sample period	$RV^{1\text{day}}$ Full	$RV^{5\text{ min}}$ Full	RV^{Mean} Full
$RV^{1\text{ sec}}$	✓	–	×
$RV^{1\text{ min}}$	✓	✓	✓
$RV^{1\text{day}}$	★	×	×
$RV^{AC1,1\text{ min}}$	✓	✓	✓
$TSRV^{tick}$	✓	–	×
$MSRV^{tick}$	✓	–	×
RK^{TH2}	✓	✓	×
RRV	✓	✓	×
$BPV^{1\text{ min}}$	✓	×	×
QRV	✓	✓	–
MedRV	✓	✓	×
RV^{Mean}	✓	✓	★
$RV^{Geo-mean}$	✓	✓	×
RV^{Median}	✓	✓	×

Optimal combinations of RV estimators

- In the paper we present estimated optimal linear combination weights across the 32 individual estimators, but no clear patterns emerge (unsurprising given multicollinearity)
- The estimated optimal combinations can also be used to test the optimality of the equally-weighted average:

$$\begin{aligned} & H_0 & : \quad & w_0^* = 0 \cap w_1^* = \ldots = w_n^* = 1/n \\ \text{vs.} \quad & H_a & : \quad & w_0^* \neq 0 \cup w_i^* \neq 1/n \text{ for some } i = 1, 2, \ldots, n \end{aligned}$$

⇒ This null is rejected with a p-value of less than 0.001

- Thus while a simple mean does well, it is possible to construct more accurate combination estimators.

Encompassing of RV estimators

- We can also test whether a single estimator "encompasses" all others, in the same spirit as Chong and Hendry (1986) and Fair and Shiller (1990):

$$\begin{aligned} H_0^i &: w_i^* = 1 \cap w_j^* = 0 \ \forall\ j \neq i \\ \text{vs.} \quad H_a^i &: w_i^* \neq 1 \cup w_j^* \neq 0 \text{ for some } j \neq i \end{aligned}$$

$\Rightarrow$ This null is rejected *for every single estimator*, with p-values all less than 0.001.

- This is very strong evidence for considering combination RV estimators: no single estimator dominates all others.

Out-of-sample comparisons of RV estimators

- We next consider comparing each of these estimators via a standard forecast experiment.
- We use the HAR model of Corsi (2004):

$$\tilde{\theta}_t = \beta_{0i} + \beta_{Di} X_{it-1} + \beta_{Wi} \frac{1}{5} \sum_{j=1}^{5} X_{i,t-j} + \beta_{Mi} \frac{1}{22} \sum_{j=1}^{22} X_{i,t-j} + \varepsilon_{it}$$

- We use Jan 1996 - Dec 1999 as the initial estimation period, and then re-estimate each day using a rolling window of 1011 days.
  - We again use $RV^{5\,min}$ as the volatility proxy
- This is then a standard volatility forecasting problem, and we can compare the forecasts using existing methods.

Out-of-sample performance of a sub-set of the estimators

	Avg. ΔQLIKE	Rank Full	Rank 00-03	Rank 04-08	In MCS? Full
$RV^{1\,sec}$	0.006	33	29	33	–
$RV^{1\,min}$	-0.012	3	5	6	✓
$RV^{AC1,1\,min}$	-0.013	1	3	9	✓
$TSRV^{tick}$	-0.002	17	19	13	–
$MSRV^{tick}$	0.001	23	31	14	–
RK^{TH2}	0.003	30	33	25	–
RRV	-0.011	7	9	7	✓
$BPV^{1\,min}$	-0.007	10	6	15	–
QRV	-0.007	13	1	31	–
MedRV	-0.007	12	2	28	–
RV^{Mean}	-0.012	4	10	2	✓
$RV^{Geo-mean}$	-0.013	2	11	1	✓
RV^{Median}	-0.011	6	12	4	–
$FCAST^{Mean}$	-0.006	15	15	10	–
$FCAST^{Geo-mean}$	-0.007	14	14	8	–
$FCAST^{Median}$	-0.005	16	16	11	–

Out-of-sample Romano-Wolf tests

Benchmark:	$RV^{1\text{day}}$	$RV^{5\text{ min}}$	RV^{Mean}	$FCAST^{Mean}$
Sample:	Full	Full	Full	Full
$RV^{1\text{ sec}}$	✓	–	×	×
$RV^{1\text{ min}}$	✓	✓	–	✓
$RV^{AC1,1\text{ min}}$	✓	✓	–	✓
$TSRV^{tick}$	✓	–	×	–
$MSRV^{tick}$	✓	–	×	×
RK^{TH2}	✓	–	×	×
RRV	✓	✓	–	✓
$BPV^{1\text{ min}}$	✓	–	–	–
QRV	✓	–	–	–
MedRV	✓	–	–	–
RV^{Mean}	✓	✓	★	✓
$RV^{Geo-mean}$	✓	✓	–	✓
RV^{Median}	✓	✓	–	✓
$FCAST^{Mean}$	✓	✓	×	★
$FCAST^{Geo-mean}$	✓	✓	×	✓
$FCAST^{Median}$	✓	✓	×	–

Combine estimators or combine forecasts?

- An interesting question arises on whether it is better to use a combination RV estimator in the HAR model and then forecast, or to estimate HAR models on individual RV estimators and then combine the forecasts.
- We compare HAR forecasts using RV^{Mean}, $RV^{Geo-mean}$ and RV^{Median} with combination forecasts $FCAST^{Mean}$, $FCAST^{Geo-mean}$, and $FCAST^{Median}$ using a simple Diebold-Mariano (1995) test

$$\begin{array}{rrcl} \text{Mean} & DM\ t\text{-}stat & = & 7.66 \\ \text{Geo-mean} & DM\ t\text{-}stat & = & 7.10 \\ \text{Median} & DM\ t\text{-}stat & = & 6.49 \end{array}$$

- Thus we find strong evidence that estimating a single HAR model on a combination RV estimator dominates using a forecast combination based on many individual forecasts.

Conclusion and summary of results

- This paper's main question: *Do combinations of RV estimators offer gains in accuracy relative to individual estimators?* $\longrightarrow$ **Yes!**

- Using a new method for comparing RV estimator accuracy and a standard out-of-sample forecast experiment, we find that combination RV estimators significantly outperform individual estimators.

  - In-sample, only two estimators ($\mathrm{RV}^{1\,\mathrm{min}}$ and $\mathrm{RV}^{AC1\,\mathrm{min}}$) significantly out-perform a simple equally-weighted average RV estimator.
  - Out-of-sample, no estimator significantly out-performs a simple equally-weighted average RV estimator.

- Further, no single RV estimator encompassed the information available in all other estimators, providing additional support for combination realised measures.