Journal of Econometrics 147 (2008) 34–46

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Realized volatility forecasting and option pricing☆

Federico M. Bandi *, Jeffrey R. Russell, Chen Yang

Graduate School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, IL, 60637, USA

ARTICLE INFO

Article history:
Available online 9 September 2008

JEL classification:
C53
G13

Keywords:
Option pricing
Microstructure noise
Volatility forecasting
Economic metrics
Realized variance
Two-scale estimator
Realized kernels

ABSTRACT

A growing literature advocates the use of microstructure noise-contaminated high-frequency data for the purpose of volatility estimation. This paper evaluates and compares the quality of several recently-proposed estimators in the context of a relevant economic metric, i.e., profits from option pricing and trading. Using forecasts obtained by virtue of alternative volatility estimates, agents price short-term options on the S&P 500 index before trading with each other at average prices. The agents' average profits and the Sharpe ratios of the profits constitute the criteria used to evaluate alternative volatility estimates and the corresponding forecasts. For our data, we find that estimators with superior finite sample Mean-squared-error properties generate higher average profits and higher Sharpe ratios, in general. We confirm that, even from a forecasting standpoint, there is scope for optimizing the finite sample properties of alternative volatility estimators as advocated by Bandi and Russell [Bandi, F.M., Russell, J.R., 2005. Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations. Working Paper; Bandi, F.M., Russell, J.R., 2008b. Microstructure noise, realized variance, and optimal sampling. Review of Economic Studies 75, 339–369] in recent work.

1. Introduction

The recent, stimulating work on nonparametric volatility estimation by virtue of high-frequency asset price data has largely focused on designing estimators with satisfactory statistical properties when realistic market microstructure contaminations play a role. Important theoretical emphasis is generally placed on the asymptotic features of the proposed estimators (see, e.g., Barndorff-Nielsen et al. (2005), Barndorff-Nielsen et al. (in press), Kalnina and Linton (2008), Zhang et al. (2005) and Zhang (2006)). A related strand of this literature focuses on the estimators' finite sample performance and the importance of optimizing this performance explicitly (Bandi and Russell, 2003, 2005, 2008b). Bandi and Russell (2008a), Barndorff-Nielsen and Shephard (2007), and McAleer and Medeiros (2008) provide thorough reviews of this growing literature.

Somewhat sparse work examines the forecasting ability of alternative high-frequency volatility estimators. This paper considers this issue and evaluates volatility forecasting from the vantage point of a relevant economic criterion. Specifically, we evaluate the profits/losses that option dealers would derive from trading on the basis of alternative volatility forecasts. To this extent, we employ a methodology proposed by Engle et al. (1990) and effectively operate in the context of an artificial ("hypothetical," in their terminology) option market. Consider two generic option traders: Trader A and Trader B. Trader A (B) estimates volatility using Method A (B). Alternative volatility estimates yield alternative volatility forecasts and, hence, different option prices. The pair-wise trades are conducted at the mid-point of the traders' prices. The trader with the highest volatility (option price) forecast will buy a straddle (a call and a put option) from his/her counterpart. All options are executed on the following day. The average dollar profits and Sharpe ratios obtained from repeated implementations of this strategy represent the economic metrics used to evaluate alternative variance estimates/forecasts. The logic is simple. If the high volatility forecast is accurate, the straddle (whose price is twice the mid-point of the call/put price forecasts in our framework) is underpriced. The trader who buys the straddle is expected to make a profit. Since a straddle is just a portfolio containing a call option and a put option, high volatility is expected to result in either option being substantially in the money. If the cost of this position is relatively low, as determined by the above-mentioned underpricing, then a profit will arise.

Traders price and trade 6-hour and 1-day (at-the-money) options on the S&P 500 index. They use high-frequency data on the Standard and Poor's depository receipts (SPIDERS) to construct the variance estimates/forecasts. We work with a variety of variance estimates/forecasts (Methods). Each trader uses a

☆ We are grateful to Loriano Mancini, Valeri Voev, two anonymous referees, and the Editor (Jiti Gao) for their useful comments. We thank the William S. Fishman Faculty Research Fund at the Graduate School of Business of the University of Chicago (Bandi), NYU Stern (Russell), and the Graduate School of Business of the University of Chicago (Bandi, Russell, and Yang) for financial support.

* Corresponding author. Tel.: +1 773 834 4352.
E-mail addresses: federico.bandi@chicagogsb.edu (F.M. Bandi), jeffrey.russell@chicagogsb.edu (J.R. Russell), cyang1@chicagogsb.edu (C. Yang).

0304-4076/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2008.09.002

method and trades with every other trader. The pair-wise trades always occur at the mid-point of the price forecasts derived from the corresponding pair of Methods. For each trader, we report the Mean of the average profits/losses and the corresponding Sharpe ratios across the pair-wise trades. We employ the following Methods:

1. Realized variance constructed using 5-min returns.
2. Realized variance constructed using 15-min returns.
3. Optimally-sampled realized variance as studied by Bandi and Russell (2003, 2008b).
4. The "near consistent" Bartlett kernel estimator in Barndorff-Nielsen et al. (2005) with an optimal (in a finite sample) choice for the number of autocovariances as suggested by Bandi and Russell (2005).
5. The two-scale estimator of Zhang et al. (2005) with an asymptotically optimal choice for the number of subsamples as suggested by Zhang et al. (2005).
6. The two-scale estimator of Zhang et al. (2005) with an optimal (in a finite sample) choice for the number of subsamples as suggested by Bandi and Russell (2005).
7. The bias-corrected two-scale estimator of Zhang et al. (2005) with an asymptotically optimal choice for the number of subsamples as suggested by Zhang et al. (2005).
8. The bias-corrected two-scale estimator of Zhang et al. (2005) with an optimal (in a finite sample) choice for the number of subsamples as suggested by Bandi and Russell (2005).
9. The flat-top *Bartlett* kernel estimator of Barndorff-Nielsen et al. (in press) with an asymptotically optimal choice for the number of autocovariances as suggested by Barndorff-Nielsen et al. (in press).
10. The flat-top *Bartlett* kernel estimator of Barndorff-Nielsen et al. (in press) with an optimal (in a finite sample) choice for the number of autocovariances as suggested by Bandi and Russell (2005).
11. The flat-top *cubic* kernel estimator of Barndorff-Nielsen et al. (in press) with an asymptotically optimal choice for the number of autocovariances as suggested by Barndorff-Nielsen et al. (in press).
12. The flat-top *cubic* kernel estimator of Barndorff-Nielsen et al. (in press) with an optimal (in a finite sample) choice for the number of autocovariances as suggested by Bandi and Russell (2005).
13. The flat-top *modified Tukey-Hanning* kernel estimator of Barndorff-Nielsen et al. (in press) with an asymptotically optimal choice for the number of autocovariances as suggested by Barndorff-Nielsen et al. (in press).
14. The flat-top *modified Tukey-Hanning* kernel estimator of Barndorff-Nielsen et al. (in press) with an optimal (in a finite sample) choice for the number of autocovariances as suggested by Bandi and Russell (2005).

Our results suggest that explicit optimization of the finite sample Mean-squared-error (MSE) properties of the proposed estimators, as advocated by Bandi and Russell (2003, 2005, 2008b), results in important economic gains. We find that the optimized (in a finite sample) flat-top kernel estimators largely constitute the most favorable Methods, both in terms of average profits and in terms of Sharpe ratios. The "near consistent" Bartlett kernel estimator and the (bias-corrected and unadjusted) two-scale estimator of Zhang et al. (2005) can perform very well, and sometimes as well as the flat-top kernel estimators, when the number of subsamples/autocovariances is carefully chosen using finite sample Methods. Choices of subsamples/autocovariances based on asymptotic, rather than finite sample, criteria lead to unnecessarily suboptimal performance. This is particularly evident in the case of the unadjusted two-scale estimator due to a potentially large finite sample (downward) bias. Optimally-sampled realized variance almost always dominates 5- and 15-min realized variance.

As mentioned, volatility forecasting in the presence of market microstructure noise is still a largely underexplored subject. Bandi and Russell (2006) and Bandi et al. (2008) use reduced-form models to show that optimally-sampled (for each trading day) realized variances (covariances) outperform realized variances (covariances) constructed using ad-hoc (5- or 15-min, say) intervals in predicting variances (covariances) out-of-sample. Ghysels and Sinko (2006a) employ the MIDAS approach of Ghysels et al. (2006) to evaluate the relative performance of realized variance based on fixed intervals, bias-corrected realized variance, and power variation. Power variation is the preferred estimator in their framework. Large (2006) uses the HAR-RV model of Corsi (2003) to find that his "alternation estimator" can have better forecasting properties than realized variance constructed using ad-hoc, fixed intervals. Aït-Sahalia and Mancini (2008) study the forecasting performance of the two-scale estimator and realized variance using a variety of simulation set-ups for stochastic volatility and microstructure noise. An empirical comparison relying on coefficients of determination from Mincer–Zarnowitz-style regressions is also provided. In their framework, the two-scale estimator outperforms realized variance both in terms of MSE and in terms of forecasting ability. Conditional predictive densities and confidence intervals for alternative integrated variance estimators are studied in Corradi et al. (2006).

Two economic metrics have been proposed, thus far. Bandi and Russell (2005, 2006) consider a *portfolio choice* problem and the long-run utility that a Mean-variance representative investor derives from alternative variance forecasts as the relevant performance criterion. A similar portfolio-based approach has been recently implemented by Bandi et al. (2008) and De Pooter et al. (2008) in a multivariate context (see Fleming et al. (2001, 2003) and West et al. (1993), in the no noise case). Bandi and Russell (2008b) study volatility forecasting for the purpose of *option pricing* as in the current paper. However, their focus is only on optimally-sampled realized variance and fixed-interval realized variance.

The extant approaches focus on subsets of the existing estimators. Their findings generally point to the preferability of optimally-sampled realized variance over realized variance constructed using fixed intervals, as well as to the out-of-sample usefulness of the consistent estimators, such as the two-scale estimator, when optimized using finite sample Methods. Stimulating recent studies towards a more comprehensive analysis of the forecasting performance of alternative volatility estimation Methods have been conducted by Andersen et al. (2006) and Ghysels and Sinko (2006b). These papers use empirically-relevant stochastic volatility models and analytic expressions to evaluate forecasting performance in linear regressions of integrated variance on alternative variance estimates. Mincer–Zarnowitz-style regression models (Andersen et al., 2006) and MIDAS regressions (Ghysels and Sinko, 2006b) are used to predict variance in practice. In Andersen et al. (2006) and Ghysels and Sinko (2006b), the forecasting metric has a statistical nature. The current paper's goal is to provide a rich comparison between alternative estimation Methods in the context of an important economic criterion. We show that estimators with superior finite sample MSE properties generally have superior predictive ability in our set-up.

The paper proceeds as follows. Section 2 describes the Methods. In Section 3 we discuss the pricing and trading mechanism. Section 4 is about the data. Section 5 contains the profit-based rankings.

2. The Methods

Consider, for simplicity, a trading day of length $h = 1$. Assume availability of $M+1$ equispaced logarithmic asset prices over [0, 1] and write

$$p_{j\delta} = p^*_{j\delta} + \eta_{j\delta}$$

or, in terms of continuously-compounded returns,

$$\underbrace{p_{j\delta} - p_{(j-1)\delta}}_{r_{j\delta}} = \underbrace{p^*_{j\delta} - p^*_{(j-1)\delta}}_{r^*_{j\delta}} + \underbrace{\eta_{j\delta} - \eta_{(j-1)\delta}}_{\varepsilon_{j\delta}},$$

where p^* denotes the *unobservable* equilibrium price, η denotes *unobservable* market microstructure noise, and $\delta = \frac{1}{M}$ represents the time distance between adjacent price observations. Assume the equilibrium price process evolves in time as a stochastic volatility local martingale, i.e.,

$$p^*_t = \int_0^t \sigma_s \mathrm{d}W_s,$$

where σ_t is a càdlàg stochastic volatility process and W_t is a standard Brownian motion.[1]

The object of econometric interest is $V = \int_0^1 \sigma_s^2 \mathrm{d}s$. The Methods attempt to identify V by mitigating the impact that the presence of the noise component η has on nonparametric volatility estimates constructed using intra-daily, noise-contaminated asset returns r. Because we use SPIDERS high-frequency mid-quotes in this study, the price formation mechanism should be understood as the SPIDERS' mid-quote formation mechanism.

In what follows, we use the symbol $\perp\!\!\!\perp$ to signify "independence." The symbol $\Rightarrow$ denotes "weak convergence."

- *Realized variance constructed using* 5-min *returns*, $\widehat{V}^{5\,\text{min}}$

The classical realized variance estimator of Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2002) is simply defined as the sum of the squared returns over the trading day, i.e.,

$$\widehat{V} = \sum_{j=1}^{M} r^2_{j\delta}.$$

In the absence of noise, $\widehat{V}$ is consistent for V as the sampling frequency increases, i.e., as $M \to \infty$ over [0, 1] (see, e.g., Protter (1995)). In the presence of empirically-reasonable noise contaminations, $\widehat{V}$ is inconsistent as shown by Bandi and Russell (2003, 2008b) and Zhang et al. (2005). In order to reduce the impact of the noise component, Andersen et al. (2001, 2003), among others, suggest using 5-min returns, rather than returns sampled at the highest frequencies, to estimate V in practice. Thus, $\widehat{V}^{5\,\text{min}}$ is equal to $\widehat{V}$ with $M = 72$ (=$6 \times 60/5$).

- *Realized variance constructed using* 15-min *returns*, $\widehat{V}^{15\,\text{min}}$

Since, provided $\eta \perp\!\!\!\perp p^*$, the bias of the realized variance estimator is equal to $M\mathbf{E}_M(\varepsilon^2)$, this bias can be further reduced by selecting a lower number of intra-daily returns for the purpose of variance estimation. Relying on plots of realized variance versus alternative sampling frequencies (i.e., "volatility signature plots") and the leveling off of realized variance around the 15-min frequency, Andersen et al. (1999, 2000) recommend using 15- or 20-min frequencies for the purpose of constructing realized variance. Hence, $\widehat{V}^{15\,\text{min}}$ is equal to $\widehat{V}$ with $M = 24(= 6 \times 60/15)$.

- *Optimally-sampled realized variance*, $\widehat{V}^{Opt}$

While low sampling frequencies reduce the extent of the estimator's bias component, they increase its variance. An optimal sampling frequency can then be chosen by optimally trading-off bias and variance as suggested by Bandi and Russell (2003, 2008b). Bandi and Russell (2003, 2008b) discuss selection of the optimal MSE-based frequency $\frac{1}{M^{Opt}}$ in the presence of noise dependence but provide a particularly simple rule-of-thumb to select this frequency when $\eta \perp\!\!\!\perp p^*$ and η is, as an approximation at least, i.i.d. in discrete time. Leverage effects are also ruled out (i.e., $W \perp\!\!\!\perp \sigma$).[2] In this case,

$$M^{Opt} \approx \left(\frac{\int_0^1 \sigma_s^4 \mathrm{d}s}{\left(\mathbf{E}(\varepsilon^2)\right)^2}\right)^{\frac{1}{3}}.$$

Clearly, M^{Opt} depends on a signal-to-noise ratio. The larger the signal coming from the equilibrium price process $Q = \int_0^1 \sigma_s^4 \mathrm{d}s$ relative to the noise component $\left(\mathbf{E}(\varepsilon^2)\right)^2$, the larger M^{Opt}. Both quantities in M^{Opt} can be evaluated. The quarticity Q can be estimated inconsistently, but with little bias, by calculating $\widehat{Q} = \frac{M}{3}\sum_{j=1}^{M} r^4_{j\delta}$ (realized quarticity) with low frequency returns, e.g., $M = 24$ or 15-min intervals. Under the same assumptions leading to M^{Opt}, the noise second moment can be estimated consistently by computing $\widehat{\mathbf{E}}(\varepsilon^2) = \sum_{j=1}^{M} r^2_{j\delta}/M$ at the highest frequency at which the data arrives (see, e.g., Bandi and Russell (2003) and Zhang et al. (2005)).[3] Hence, $\widehat{V}^{Opt} = \widehat{V}$ with $\widehat{M}^{opt} = \left(\widehat{Q}/\left(\widehat{\mathbf{E}}(\varepsilon^2)\right)^2\right)^{1/3}$.

- *The "near consistent" Bartlett kernel estimator in* Barndorff-Nielsen et al. (2005) *with an optimal (in a finite sample) choice for the number of autocovariances as suggested by* Bandi and Russell (2005).

Here, and below, we assume the data is not necessarily equispaced. Consider a HAC-type estimator defined as

$$\widehat{V}^{Bar} = \left(\frac{M-1}{M}\frac{q-1}{q}\right)\widehat{\gamma}_0 + 2\sum_{s=1}^{q}\left(\frac{q-s}{q}\right)\widehat{\gamma}_s,$$

where $\widehat{\gamma}_s = \sum_{j=1}^{M-s} r_j r_{j+s}$ is the s-th return (realized) autocovariance. Clearly, $\widehat{\gamma}_0 = \widehat{V}$, hence the estimator weighs the classical realized variance estimator (possibly applied to non-equispaced data) and the first q return autocovariances by virtue of Bartlett-type kernel weights. This estimator is in the spirit of the first-order bias-corrected estimator studied by Zhou (1996), Hansen and Lunde (2006), and Oomen (2005).

Under i.i.d. η's and $p^* \perp\!\!\!\perp \eta$, Barndorff-Nielsen et al. (2005) show that, if $M, q \to \infty$ with $\frac{q}{M} \to 0$ and $\frac{q^2}{M} \to \infty$, even though inconsistent for V, $\widehat{V}^{Bar}$ has a small (empirically) limiting variance given by $4\left(\mathbf{E}(\eta^2)\right)^2$.

We implement $\widehat{V}^{Bar}$ by selecting the number of autocovariances on the basis of the MSE-based approximation suggested by

[1] For thorough discussions of this model and the economics of the price contaminations η we refer the reader to the review papers of Bandi and Russell (2008a), Barndorff-Nielsen and Shephard (2007), and McAleer and Medeiros (2008).

[2] Bandi and Russell (2008b) discuss the validity of this assumption at length.

[3] Only the i.i.d. noise assumption is important for consistent estimation of the second moment of the noise by virtue of the estimator described in the text. One can allow for dependence between the noise and the equilibrium price as well as for leverage effects with no impact on asymptotic inference (Bandi and Russell, 2008b). Bandi and Russell (2008a) discuss extensions to the dependent noise case.

Bandi and Russell (2005)[4]:

$$q^{BR} \approx \left(\frac{3}{2} \frac{\frac{V^2}{M^2}}{Q} \right)^{1/3} M.$$

As earlier in the case of M^{Opt}, q^{BR} is determined on the basis of a bias-variance trade-off. Leaving proportionality factors aside, the term $\frac{V^2}{M^2}$ represents the leading term of the estimator's (squared) bias, whereas Q represents the leading term of its variance. The larger the bias term relative to the variance term, the larger the number of autocovariances. As earlier, preliminary, roughly unbiased, estimates of V and Q can be obtained by employing realized variance, $\widehat{V}$, and realized quarticity, $\widehat{Q}$, with low frequency returns (for instance, 15-min returns).

- *The two-scale estimator of* Zhang et al. (2005) *with an asymptotically optimal choice for the number of subsamples as suggested by* Zhang et al. (2005), $\widehat{V}^{ZMA}_{ZMA}$

Consider q non-overlapping sub-grids $\Psi^{(i)}$ of the original grid of M arrival times with $i = 1, \ldots, q$. The first sub-grid starts at t_0 and takes every q-th arrival time, i.e., $\Psi^{(1)} = (t_0, t_{0+q}, t_{0+2q}, \ldots,)$, the second sub-grid starts at t_1 and also takes every q-th arrival time, i.e., $\Psi^{(2)} = (t_1, t_{1+q}, t_{1+2q}, \ldots,)$, and so on. Given the generic i-th sub-grid of arrival times, the corresponding realized variance estimator is defined as

$$\widehat{V}^{(i)} = \sum_{t_j, t_{j+} \in \Psi^{(i)}} \left(p_{t_{j+}} - p_{t_j} \right)^2, \tag{1}$$

where t_j and t_{j+} denote adjacent elements in $\Psi^{(i)}$. Zhang et al's two-scale or subsampling estimator is constructed as

$$\widehat{V}^{ZMA} = \frac{\sum_{i=1}^{q} \widehat{V}^{(i)}}{q} - \overline{M}\widehat{\mathbf{E}}(\varepsilon^2), \tag{2}$$

where $\overline{M} = \frac{M-q+1}{q}$, $\widehat{\mathbf{E}}(\varepsilon^2) = \frac{\sum_{j=1}^{M} \left(p_{t_{j+}} - p_{t_j} \right)^2}{M}$ is an estimate of the second moment of the noise return, as discussed earlier, and $\overline{M}\widehat{\mathbf{E}}(\varepsilon^2)$ is a bias-correction. Thus, the estimator averages the realized variance estimates constructed on the basis of subsamples and bias-corrects them.

Under i.i.d. η's and $p^* \perp\!\!\!\perp \eta$, Zhang et al. (2005) show that, if $M, q \to \infty$ with $\frac{q}{M} \to 0$ and $\frac{q^2}{M} \to \infty$, the estimator is consistent. Specifically, if $q = cM^{2/3}$,

$$M^{1/6} \left(\widehat{V}^{ZMA} - V \right) \Rightarrow \left(\sqrt{8c^{-2} \left(\mathbf{E}(\eta^2) \right)^2 + c \frac{4}{3} Q} \right) N(0, 1). \tag{3}$$

The constant c can be selected optimally in order to minimize the estimator's limiting variance. This minimization lead to an asymptotically optimal number of subsamples given by

$$\begin{aligned} q^{ZMA} &= c^{ZMA} M^{2/3} = \left(\frac{16 \left(\mathbf{E}(\eta^2) \right)^2}{\frac{4}{3} Q} \right)^{1/3} M^{2/3} \\ &= \left(\frac{3 \left(\mathbf{E}(\varepsilon^2) \right)^2}{Q} \right)^{1/3} M^{2/3} \end{aligned} \tag{4}$$

(Zhang et al., 2005). Q and $\mathbf{E}(\varepsilon^2)$ can be estimated as in the case of M^{opt}. In sum, $\widehat{V}^{ZMA}_{ZMA} = \widehat{V}^{ZMA}$ with $q = q^{ZMA}$.

- *The two-scale estimator of* Zhang et al. (2005), *with an optimal (in a finite sample) choice for the number of subsamples as suggested by* Bandi and Russell (2005), $\widehat{V}^{ZMA}_{BR}$

Barndorff-Nielsen et al. (2005, 2007) have shown that the two-scale estimator can be rewritten as a modified Bartlett-type kernel estimator, i.e.,

$$\widehat{V}^{ZMA} = \left(1 - \frac{M-q+1}{Mq} \right) \widehat{\gamma}_0 + 2 \sum_{s=1}^{q} \left(\frac{q-s}{q} \right) \widehat{\gamma}_s - \frac{1}{q} \vartheta_q, \tag{5}$$

where $\widehat{\gamma}_s = \sum_{j=1}^{M-s} r_j r_{j+s}$, $\vartheta_1 = 0$, and $\vartheta_q = \vartheta_{q-1} + (r_1 + \cdots + r_{q-1})^2 + (r_{M-q+2} + \cdots + r_M)^2$ for $q \geq 2$. The term $\frac{1}{q}\vartheta_q$, which mechanically derives from subsampling, is crucial for the estimator's consistency and differentiates $\widehat{V}^{ZMA}$ from the inconsistent Bartlett kernel estimator $\widehat{V}^{Bar}$.

Under i.i.d. η's, $p^* \perp\!\!\!\perp \eta$, and $W \perp\!\!\!\perp \sigma$, Bandi and Russell (2005) find that, despite the inconsistency of $\widehat{V}^{Bar}$, the finite sample MSEs of $\widehat{V}^{ZMA}$ and $\widehat{V}^{Bar}$ are very similar in practice.[5] They suggest using the same rule-of-thumb q^{BR} as for $\widehat{V}^{Bar}$ when constructing $\widehat{V}^{ZMA}$. Hence, $\widehat{V}^{ZMA}_{BR}$ is equal to $\widehat{V}^{ZMA}$ with $q = q^{BR}$.

- *The bias-corrected two-scale estimator of* Zhang et al. (2005) *with an asymptotically optimal choice for the number of subsamples as suggested by* Zhang et al. (2005), $\widehat{V}^{ZMAadj}_{ZMA}$

Albeit consistent, the two-scale estimator is (downward) biased in finite samples. However, under i.i.d. η's, $p^* \perp\!\!\!\perp \eta$, and $W \perp\!\!\!\perp \sigma$, the form of the bias has been provided (Bandi and Russell, 2005; Barndorff-Nielsen et al., 2005).[6] Therefore, following a suggestion contained in Zhang et al. (2005), the estimator can be bias-corrected. The bias-corrected estimator can be written as $\widehat{V}^{ZMAadj} = c(q, M)\widehat{V}^{ZMA}$, where

$$c(q, M) = \left(\frac{qM - 1 + 2q - q^2 - M}{qM} \right)^{-1}.$$

Note that, for reasonable values of q and M, the factor $c(q, M)$ will increase the magnitude of the estimates, thereby off-setting the downward bias of the original estimator.[7] Because $\widehat{V}^{ZMAadj}$

[4] This approximation is valid under the same conditions yielding "near consistency" of the estimator as in Barndorff-Nielsen et al. (2005) and $W \perp\!\!\!\perp \sigma$. The approximation relies on the finite sample MSE of the estimator. The full MSE can be optimized directly (Bandi and Russell, 2005).

[5] When studying $\widehat{V}^{ZMA}$, their results also rely on the assumption $\sigma_j^2 = \int_{t_{j-1}}^{t_j} \sigma_s^2 ds = \frac{V}{M} \forall j$. This condition is exact when sampling occurs in business time (see, e.g., Oomen (2005)). It is an approximation when volatility does not vary drastically during the day. Importantly, the condition is solely introduced to facilitate empirical tractability of the resulting MSE expansion. In other words, the exact MSE expansion of $\widehat{V}^{ZMA}$ may be easily provided while dispensing with this assumption. As an example, below we report the general form of the estimator's bias.

[6] The exact bias is given by

$$\left(\frac{-M+q-1}{qM} \right) V - \frac{1}{q} \sum_{s=1}^{q-1} (q-s)(\sigma_s^2 + \sigma_{M+1-s}^2).$$

Under the additional assumption $\sigma_j^2 = \int_{t_{j-1}}^{t_j} \sigma_s^2 ds = \frac{V}{M} \forall j$ (discussed above), the bias reduces to

$$\left(\frac{-M+2q-q^2-1}{qM} \right) V.$$

[7] The form of the bias-correction is slightly different from that contained in Zhang et al. (2005) where $\widehat{V}^{ZMAadj} = \left(\frac{qM-1+q-M}{qM} \right)^{-1} \widehat{V}^{ZMA}$. Specifically, under $\sigma_j^2 = \int_{t_{j-1}}^{t_j} \sigma_s^2 ds = \frac{V}{M} \forall j$, our bias-correction is exact. Hence, the estimator's finite sample MSE reduces to its finite sample variance in Eq. (6). Zhang et al. (2005) do not adjust for the term $\frac{1}{q} \sum_{s=1}^{q-1} (q-s)(\sigma_s^2 + \sigma_{M+1-s}^2)$ (or $\left(\frac{q^2-q}{qM} \right) V$ under $\sigma_j^2 = \int_{t_{j-1}}^{t_j} \sigma_s^2 ds = \frac{V}{M} \forall j$) in the bias expansion. Since this term is of order $\frac{q}{M} - \frac{1}{M}$, the difference between the exact correction discussed here and the approximate correction in Zhang et al. (2005) is asymptotically immaterial and can be empirically small.

is asymptotically equivalent to $\widehat{V}^{ZMA}$, the asymptotically optimal number of subsamples is given by q^{ZMA}. This choice lead to $\widehat{V}^{ZMAadj}_{ZMA}$.

- *The bias-corrected two-scale estimator of* Zhang et al. (2005) *with an optimal (in a finite sample) choice for the number of subsamples as suggested by* Bandi and Russell (2005), $\widehat{V}^{ZMAadj}_{BR}$.

The optimal (in a finite sample) number of subsamples of $\widehat{V}^{ZMAadj}$ is provided by optimizing the finite sample variance of the estimator (Bandi and Russell, 2005). Under i.i.d. η's, $p^* \perp\!\!\!\perp \eta$, $W \perp\!\!\!\perp \sigma$, and $\sigma_j^2 = \int_{t_{j-1}}^{t_j} \sigma_s^2 ds = \frac{V}{M} \forall j$, write

$$\mathrm{var}\left(\widehat{V}^{ZMAadj}\right) = (c(q, M))^2 \,\mathrm{var}\left(\widehat{V}^{ZMA}\right),$$

where, if $\frac{q}{M} \le 1/2$,

$$\begin{aligned}
\mathrm{var}\left(\widehat{V}^{ZMA}\right) &= K^{ZMA} - \frac{1}{3}(Q + V^2)\left(\frac{q}{M}\right)^2 \\
&+ \left(-\frac{1}{3}V^2\frac{1}{M} - 4V^2\frac{1}{M^2} + \frac{4}{3}Q\right)\frac{q}{M} \\
&+ \left[-\frac{4}{M^4}\left(Q + V^2\right) + \left(\frac{8\sigma_\eta^4 + 16\sigma_\eta^2 V - 8Q - \frac{56}{3}V^2}{M^3}\right)\right. \\
&\left. + \left(\frac{24\sigma_\eta^2 V - \frac{10}{3}Q + 8\sigma_\eta^4}{M^2}\right) + \left(\frac{-8\sigma_\eta^4 + 8\sigma_\eta^2 V}{M}\right)\right]\frac{M}{q} \\
&+ \left[\frac{2}{M^5}Q + \left(\frac{-4\sigma_\eta^4 - 8\sigma_\eta^2 V + 4Q - 8V^2}{M^4}\right)\right. \\
&+ \left(\frac{-4\sigma_\eta^4 - 16\sigma_\eta^2 V + 2Q}{M^3}\right) \\
&\left. + \left(\frac{8\sigma_\eta^4 - 8\sigma_\eta^2 V}{M^2}\right) + \frac{8}{M}\sigma_\eta^4\right]\frac{M^2}{q^2},
\end{aligned} \tag{6}$$

and

$$\begin{aligned}
K^{ZMA} &= (-4\sigma_\eta^4 - 8V\sigma_\eta^2)\frac{1}{M} + \left(-4\sigma_\eta^4 - 8\sigma_\eta^2 V + \frac{13}{3}Q\right. \\
&\left. + \frac{79}{3}V^2\right)\frac{1}{M^2} + \frac{1}{M^3}\left(2Q + 8V^2\right)
\end{aligned}$$

with $\sigma_\eta^2 = \mathbf{E}(\eta^2)$. The optimal q ($=q^{BR}$) is now simply the minimizer of $\mathrm{var}\left(\widehat{V}^{ZMAadj}\right)$. We call the resulting estimator $\widehat{V}^{ZMAadj}_{BR}$.

- *The flat-top Bartlett kernel estimator of* Barndorff-Nielsen et al. (in press) *with an asymptotically optimal choice for the number of autocovariances as suggested by* Barndorff-Nielsen et al. (in press), $\widehat{V}^{BNHLS(Bar)}_{BNHLS}$

In recent work, Barndorff-Nielsen et al. (in press) have advocated using unbiased (under i.i.d. η's and $p^* \perp\!\!\!\perp \eta$) flat-top symmetric kernels of the type

$$\widehat{V}^{BNHLS} = \widehat{\gamma}_0 + \sum_{s=1}^{q} w_s(\widehat{\gamma}_s + \widehat{\gamma}_{-s}), \tag{7}$$

where $\widehat{\gamma}_s = \sum_{j=1}^{M} r_j r_{j-s}$ with $s = -q, \ldots, q$, $w_s = k\left(\frac{s-1}{q}\right)$ and $k(.)$ is a function on $[0, 1]$ satisfying $k(0) = 1$ and $k(1) = 0$. If $q = cM^{2/3}$, this family of estimators is consistent (at rate $M^{1/6}$) and asymptotically mixed normal. When $k(x) = 1 - x$ (the Bartlett case), the limiting variance of $\widehat{V}^{BNHLS}$ is the same as that of the two-scale estimator. Hence, q can be chosen asymptotically as suggested by Zhang et al. (2005).

- *The flat-top Bartlett kernel estimator of* Barndorff-Nielsen et al. (in press) *with an optimal (in a finite sample) choice for the number of autocovariances as suggested by* Bandi and Russell (2005), $\widehat{V}^{BNHLS(Bar)}_{BR}$

As earlier, Bandi and Russell (2005) provide an alternative way to choose the number of autocovariances in finite samples. Consistent with the case of $\widehat{V}^{ZMA}$ (or $\widehat{V}^{ZMAadj}$) and $\widehat{V}^{Bar}$, for a generic $k(.)$, they advocate choosing q as $\phi^{BR}M$ with $0 < \phi^{BR} \le 1$, where ϕ^{BR} minimizes the finite sample MSE of the estimator, i.e.,[8]

$$MSE = (\mathrm{bias}(\phi))^2 + \mathrm{var}(\phi),$$

where

$$\mathrm{bias}(\phi) = 0$$

and

$$\begin{aligned}
\mathrm{var}(\phi) &= \frac{Q}{M} w^\top \Omega_1 w + 4\left(\mathbf{E}(\eta^2)\right)^2 M(w^\top \Omega_2 w) \\
&+ 4\left(\mathbf{E}(\eta^2)\right)^2 (w^\top \Omega_3 w) + (2\mathbf{E}(\eta^2)V)4(w^\top \Omega_4 w),
\end{aligned}$$

with

$$w = \left(1, 1, k\left(\frac{1}{\phi M}\right), \ldots, k\left(\frac{\phi M - 1}{\phi M}\right)\right)^\top,$$

and Ω_a $a = 1, \ldots, 4$ are $(\phi M + 1, \phi M + 1)$ square matrices. For $j \le \phi M$, the matrices Ω_1 and Ω_4 are defined as follows:

$$\begin{aligned}
&\Omega_1[1,1] = 2, \quad \Omega_1[1+j, 1+j] = 4, \\
&\Omega_4[1,1] = 1, \quad \Omega_4[2,1] = -1, \quad \Omega_4[1,2] = -1, \\
&\Omega_4[2,2] = 2, \quad \Omega_4[1+j, 1+j] = 2, \\
&\Omega_4[1+j, j] = -1, \quad \Omega_4[j, j+1] = -1,
\end{aligned}$$

and 0 otherwise. For $j \le \phi M - 1$, the matrices Ω_2 and Ω_3 are defined as follows:

$$\begin{aligned}
&\Omega_2[1,1] = 3, \quad \Omega_2[1,2] = -4, \quad \Omega_2[2,1] = -4, \\
&\Omega_2[2,2] = 7, \\
&\Omega_2[2+j, 2+j] = 6, \quad \Omega_2[2+j, 1+j] = -4, \\
&\Omega_2[1+j, 2+j] = -4, \\
&\Omega_2[2+j, j] = 1, \quad \Omega_2[j, 2+j] = 1, \\
&\Omega_3[1,1] = -1, \quad \Omega_3[1,2] = 2, \quad \Omega_3[2,1] = 2, \\
&\Omega_3[2,2] = -4.5, \quad \Omega_3[j+2, j+2] = -3(j+1) - 1, \\
&\Omega_3[2+j, 1+j] = 2(j+1), \quad \Omega_3[1+j, 2+j] = 2(j+1), \\
&\Omega_3[2+j, j] = -(j+1)/2, \quad \Omega_3[j, 2+j] = -(j+1)/2,
\end{aligned}$$

and 0 otherwise. Hence, $\widehat{V}^{BNHLS(Bar)}_{BR}$ is equal to $\widehat{V}^{BNHLS}$ with $q = \phi^{BR}M$ and $k(x) = 1 - x$.

- *The flat-top cubic kernel estimator of* Barndorff-Nielsen et al. (in press) *with an asymptotically optimal choice for the number of autocovariances as suggested by* Barndorff-Nielsen et al. (in press), $\widehat{V}^{BNHLS(Cubic)}_{BNHLS}$

Consider again the class of estimators in Eq. (7). Barndorff-Nielsen et al. (in press) have shown that, if $k(.)$ is also chosen in such a way as to guarantee that $k'(0) = 0$ and $k'(1) = 0$, the number of autocovariances can be selected as $q = cM^{1/2}$ and the estimator is consistent at rate $M^{1/4}$. Specifically, in Box I,

[8] The MSE is obtained under i.i.d. η's and $p^* \perp\!\!\!\perp \eta$ (the same conditions used to prove the limiting properties of this class of estimators as well as the limiting properties of the estimators discussed above) as well as $W \perp\!\!\!\perp \sigma$. In addition, normality of the errors is assumed solely to dispense with the estimation of the fourth noise moment. This assumption can be easily relaxed (Barndorff-Nielsen et al., in press; Bandi and Russell, 2005).

$$M^{1/4}\left(\widehat{V}^{BNHLS}-V\right)\Rightarrow\left(\sqrt{4ck_\bullet^{0,0}Q-8c^{-1}k_\bullet^{0,2}\mathbf{E}(\eta^2)\left(V+\frac{\mathbf{E}(\eta^2)}{2}\right)+4\left(\mathbf{E}(\eta^2)\right)^2c^{-3}\left\{k'''(0)+k_\bullet^{0,4}\right\}}\right)N(0,1)$$

Box I.

where $k_\bullet^{0,0}=\int_0^1 k^2(x)\mathrm{d}x$, $k_\bullet^{0,2}=\int_0^1 k(x)k''(x)\mathrm{d}x$, and $k_\bullet^{0,4}=\int_0^1 k(x)k''''(x)\mathrm{d}x$. Importantly, when $k(x)=1-3x^2+2x^3$, the limiting distribution of $\widehat{V}^{BNHLS}$ is the same as that of the multi-scale estimator of Zhang (2006). We define $\widehat{V}_{BNHLS}^{BNHLS(Cubic)}$ as $\widehat{V}^{BNHLS(Cubic)}$ with $q=cM^{1/2}$ and c chosen as the minimizer of the limiting variance in Box I.

- *The flat-top cubic kernel estimator of* Barndorff-Nielsen et al. (in press) *with an optimal (in a finite sample) choice for the number of autocovariances as suggested by* Bandi and Russell (2005), $\widehat{V}_{BR}^{BNHLS(Cubic)}$

  We define $\widehat{V}_{BR}^{BNHLS(Cubic)}$ as $\widehat{V}^{BNHLS}$ with $q=\phi^{BR}M$ and $k(x)=1-3x^2+2x^3$.

- *The flat-top modified Tukey-Hanning kernel estimator of* Barndorff-Nielsen et al. (in press) *with an asymptotically optimal choice for the number of autocovariances as suggested in* Barndorff-Nielsen et al. (in press), $\widehat{V}_{BNHLS}^{BNHLS(TH)}$

  Finally, we consider the modified Tukey-Hanning flat-top kernel estimator since it is asymptotically more efficient than the cubic estimator, i.e., $k(x)=\left(1-\cos\pi(1-x)^2\right)/2$ (Barndorff-Nielsen et al., in press). We define $\widehat{V}_{BNHLS}^{BNHLS(TH)}$ as $\widehat{V}^{BNHLS(TH)}$ with $q=cM^{1/2}$ and c chosen as the minimizer of the limiting variance in Box I.

- *The flat-top modified Tukey-Hanning kernel estimator of* Barndorff-Nielsen et al. (in press) *with an optimal (in a finite sample) choice for the number of autocovariances as suggested in* Bandi and Russell (2005), $\widehat{V}_{BR}^{BNHLS(TH)}$

  We define $\widehat{V}_{BR}^{BNHLS(TH)}$ as $\widehat{V}^{BNHLS}$ with $q=\phi^{BR}M$ and $k(x)=\left(1-\cos\pi(1-x)^2\right)/2$.

2.1. Out-of-sample forecasting

All variance estimates are obtained using intra-daily high-frequency returns over a 6-hour (10 am to 4 pm) trading period. The choice variables, such as the optimal sampling frequency of the realized variance estimator, the number of subsamples of the two-scale estimators, and the number of autocovariances of the flat-top symmetric kernel estimators, for instance, should be interpreted as time-varying. Specifically, for each trading day, optimization of the (finite sample and asymptotic) MSE of the estimators translates into the selection of an optimal number of intra-daily returns in the case of realized variance and into the selection of an optimal number of autocovariances/subsamples, for a *given* number of intra-daily returns, in the case of the alternative kernel-based estimators.

The out-of-sample forecasts are derived using ARFIMA models.[9] For each method, we use 3 sample lengths to estimate the ARMA parameters: (1) 1000 days, (2) 1500 days, and (3) the entire set of daily data with no less than 1500 observations.[10] This effectively translates into a total of $3\times14=42$ Methods. As in Engle et al. (1990) we add three additional Methods, namely the average of the daily forecasts (Mean), and the daily minimum (Min) and maximum (Max) forecast, for a grand total of 45 Methods. As emphasized by Engle et al. (1990), the average of n forecasts that are conditionally independent and of similar quality will converge to an accurate forecast at a fast rate. Failure to do so indicates large economic differences and/or dependence of the forecasts. The use of the minimum (maximum) forecast should highlight the effects of persistent upward (downward) biases in the forecasts. Specifically, if the distribution of the volatility forecasts is roughly symmetric (which appears to be realistic for our data) and the forecasts are largely downward biased, then the maximum forecast should have better performance than the minimum forecast and the forecasts that are considerably downward biased. Conversely, if the forecasts are largely upward biased, then the minimum forecast should have better performance than the maximum forecast and the forecasts that are very upward biased.

To assess the influence of the overnights in variance forecasting through high-frequency asset price data, we consider two scenarios. The first involves pricing 6-hour options. Since the intra-daily variance estimates are for a 6-hour period, straightforward one-step-ahead forecasting yields the needed variance prediction. The second scenario involves pricing 1-day options. To account for the overnights (i.e., price changes between 4 pm and 10 am of the following day) we multiply each original 6-hour variance estimate $\widetilde{V}$ (before forecasting) by a constant factor ζ defined as

$$\zeta=\frac{\sum_{i=1}^{n}\left(R_i^{S\&P500}\right)^2}{\sum_{i=1}^{n}\widetilde{V}_i},\tag{8}$$

where $R_i^{S\&P500}$ is the daily return on the S&P 500 index for day i and n is the total number of days in our sample. This procedure ensures that the average of the transformed variances, i.e., $\zeta\widetilde{V}$, is equal to the average of the squared daily returns. Hence, ζ will blow up the 6-hour variance estimates.[11]

3. Option pricing and option trading

We follow Engle et al. (1990) in designing an hypothetical option market. Every agent is associated with a method. Given his/her method the agent prices 6-hour or 1-day options on a $1 share of the S&P 500 index. The exercise price of the options is taken to be $1, the risk-free rate is taken to be zero. Agents use

[9] A previous version of the paper used ARMA specifications. The resulting ranking of the forecasts was virtually the same as that reported below.

[10] The fractional d parameter is estimated by virtue of the GPH estimator (Geweke and Porter-Hudak, 1983) using the full sample. Because this parameter can hardly be estimated efficiently, we opt for using the longest span of available data for its estimation. While this choice reduces the genuine out-of-sample nature of our empirical exercise, we find that alternative, reasonable choices of the d parameter do not affect our results in any meaningful way. The d estimates are traditional and range between 0.3 for $\widehat{V}^{5\,\min}$ and 0.49 for $\widehat{V}_{ZMA,BR}^{ZMAadj}$.

[11] Hansen and Lunde (2005) provide a theoretical justification for this traditional adjustment while studying the optimal combination of overnight squared returns and intra-daily realized variance for the purpose of daily integrated variance estimation.

Black–Scholes to price. Under these assumptions, the predicted call price is given by

$$P_t = 2\Phi\left(\frac{1}{2}\sigma_t\right) - 1,$$

where Φ is the cumulative normal distribution and σ_t is a specific volatility forecast, as delivered by a method. By put/call parity, P_t is also the corresponding put price. We are now specific about the various stages of the pricing and trading process.

1. Given a method, each agent computes his/her fair Black–Scholes option price for either a 6-hour (at-the-money) option or a 1-day (at-the-money) option on a $1 share of the S&P 500 index.
2. The pair-wise trades take place. The trades are conducted at the mid-point of the agents' prices. Agents with variance forecasts higher than the mid-point will perceive the options to be underpriced. Hence, they will buy from their counterpart. Importantly, the agents buy or sell straddles (one put and one call). Hence, agents with high variance forecasts effectively speculate on volatility in that, on the one hand, they perceive the straddle to be underpriced while, on the other hand, they count on either option to end up considerably in the money due to the expected high volatility.

   Finally, the positions are hedged. The delta, or hedge ratio, of the call option is given by $\Phi\left(\frac{1}{2}\sigma_t\right)$. Hence, a trader who buys a call should go short $\Phi\left(\frac{1}{2}\sigma_t\right)$ shares of the stock for a riskless (given small changes in the stock price) position in the option and in the stock. Similarly, a trader who buys a put should go long $1-\Phi\left(\frac{1}{2}\sigma_t\right)$ for a riskless position. The hedge ratio of the straddle is the sum of the two hedge ratios, i.e., $1-2\Phi\left(\frac{1}{2}\sigma_t\right)$.

   The daily profit to a trader who buys the straddle is then given by

$$\left|R_t^{S\&P500}\right| - 2P_t + R_t^{S\&P500}\left(1 - 2\Phi\left(\frac{1}{2}\sigma_t\right)\right).$$

   The daily profit to a seller is given by

$$2P_t - \left|R_t^{S\&P500}\right| - R_t^{S\&P500}\left(1 - 2\Phi\left(\frac{1}{2}\sigma_t\right)\right).$$

3. For each trading day, the profits/losses are computed for each agent (method). The total daily profit for each agent is just the sum of the pair-wise profits divided by $l - 1$, where l is the total number of agents (Methods), i.e., 45. The profits/losses are then averaged across days.

The goal of this paper is to use a dollar metric in comparing competing volatility forecasts. In doing so, we consider a market where all agents effectively "price" volatility using the same model, a classical Black–Scholes model. The use of Black–Scholes places agents on equal footing. While our metric is sensible, future work could consider the sensitivity of our rankings (as reported below) to the assumptions made regarding the pricing model and/or the option's horizon. Because we are pricing at-the-money options over a short horizon it is likely that alternative pricing models would deliver similar rankings.

We conclude this section with two observations. First, by design, traders with median forecasts (or forecasts near the median value) have the potential to make profits off of market making (by selling to those with higher price forecasts and buying from those with lower price forecasts) even if their volatility forecasts are inaccurate. By looking at the resulting traders' positions, we find that the upward (downward) biased forecasts lead to an overwhelming majority of buy (sell) orders, as expected. Importantly, the best performing Methods have positions that are nicely dispersed around the neutral (market-making) position, with no obvious emphasis on the neutral position. Second, comparing multiple forecasts in an hypothetical market is not an obvious task. While our current set-up provides a meaningful comparison, future work should consider the case of a single market price for all transactions. Consider the price associated with the trader with the median forecast, for instance. On the one hand, differently from our current framework, traders should be allowed to choose whether to trade with other traders or not. If this were not the case, then some agents above (or below) the median price would engage in trades with a negative expected profit.[12] On the other hand, agents on different sides of the market, as determined by the clearing (median) price, should trade with each other and/or with the trader associated with the median price forecast (the market maker).

4. The data

We employ high-frequency data on the Standard and Poor's depository receipts (SPIDERS) to construct the index's intra-daily volatility estimates. SPIDERS are shares in a trust which owns stocks in the same proportion as that found in the S&P 500 index. SPIDERS trade like a stock (with the ticker symbol SPY on the Amex) at approximately one-tenth of the level of the S&P 500 index. They are widely used by institutions and traders as bets on the overall direction of the market or as a means of passive management (see, e.g., Elton et al. (2002)).

Our sample extends from January 2, 1998 to March 31, 2006. We use SPIDERS mid-quotes on the NYSE. We remove quotes whose associated price changes and/or spreads are larger than 10%. Table 1 contains descriptive statistics. In our sample, the average duration between quote updates is 11.53 seconds. The average spread and the average price level are 0.0015 and 117.27, respectively.

The second moment of the noise $\mathbf{E}(\varepsilon^2)$, V, and Q are necessary inputs in our sampling frequency and bandwidth selection procedures. We estimate $\mathbf{E}(\varepsilon^2)$ using sample second moments of quote-to-quote continuously-compounded returns. The variance and quarticity estimates are obtained by using $\widehat{V}$ (realized variance) and $\widehat{Q}$ (realized quarticity) with fixed, 15-min, calendar-time intervals.[13] The prevailing quote method is used in the absence of a quote. The out-of-sample forecasts are derived by virtue of an ARFIMA(1, 1) model. This simple specification provides a reasonable and parsimonious model for all series. The standard errors used to evaluate the statistical significance of the profits are Newey-West standard errors. The correlation structures of the various profits is mild. We use 6 autocovariances for all profit series based on their inspection. The results are insensitive to the use of an alternative, reasonable number of autocovariances.

The average optimal sampling interval of the realized variance estimator is 5.7 min. The average optimal (in a finite sample) number of autocovariances of $\widehat{V}_{BR}^{ZMAadj}$, $\widehat{V}_{BR}^{ZMA}$ (and $\widehat{V}^{Bar}$), $\widehat{V}_{BR}^{BNHLS(Bar)}$, $\widehat{V}_{BR}^{BNHLS(Cub)}$, and $\widehat{V}_{BR}^{BNHLS(TH)}$ is 7, 16, 7, 6, and 10, respectively. The average asymptotically optimal number of autocovariances of $\widehat{V}_{ZMA}^{ZMA}$ (and $\widehat{V}_{ZMA}^{ZMAadj}$), $\widehat{V}_{BNHLS}^{BNHLS(Bar)}$, $\widehat{V}_{BNHLS}^{BNHLS(Cub)}$, and $\widehat{V}_{BNHLS}^{BNHLS(TH)}$ is 5, 5, 6, and 8, respectively (see Table 1). These figures are very consistent with values obtained by Bandi and Russell (2005, 2006) using alternative assets and sample periods. Finite sample criteria lead

12 Consider two traders whose price forecasts are above the median price forecast, for example. The trader whose price forecast is relatively lower would sell even though he/she deems the straddle to be underpriced.

13 Bandi and Russell (2008b) evaluate by simulation the empirical validity of this straightforward (from an applied standpoint), albeit admittedly inefficient, method to estimate the price moments of interest. As is well-known, efficient estimation of the integrated quarticity is an open research issue.

Table 1
Summary statistics

Avg. dur.	Avg. sprd.	Avg. price	M		Opt. inter.
11.53	0.0015	117.27	3648		5.72
Asy. q		ZMA/ZMAadj	FlatBart	FlatCubic	FlatTukey
Mean		5	5	6	8
Max		185	208	129	202
Min		1	1	1	1
Std		10.22	10.99	6.99	10.85
Finite q	ZMAadj	ZMA/Bart	FlatBart	FlatCubic	FlatTukey
Mean	7	16	7	6	10
Max	216	57	185	99	189
Min	1	3	1	1	1
Std	11.18	6.45	10.18	6.53	9.19

We report summary statistics about the data (mid-quotes on SPIDERS between January 2, 1998 and March 31, 2006): average duration between mid-quote updates (in seconds), average spread, average price, and average number of daily observations. Opt. Inter. denotes the average optimal sampling interval (in minutes) of the realized variance estimator. The remaining statistics summarize the empirical distribution of the optimal (finite sample and asymptotic) number of autocovariances of the "near consistent" Bartlett kernel estimator (Bart), of the bias-corrected and unadjusted two-scale estimators (ZMAadj and ZMA), and of the flat-top kernel estimators in their Bartlett, cubic, and modified Tukey-Hanning versions (FlatBart, FlatCubic, and FlatTukey).

to a number of autocovariances which is larger on average (but generally less volatile) than asymptotic criteria. The difference is particularly evident for the unadjusted two-scale estimator. The "near consistent" Bartlett kernel estimator and the unadjusted two-scale estimator have favorable asymptotic properties but, as discussed in Bandi and Russell (2005), display a potentially large finite sample (downward) bias in general. The optimal (in a finite sample) choice of the number of autocovariances/subsamples is affected by this bias component. Specifically, effective bias reduction in a finite sample requires the selection of a fairly large number of autocovariances/subsamples (16, on average, in our case). The optimal (average) asymptotic choice (5) is likely to leave a substantial time-varying bias component in the resulting estimates. The class of flat-top symmetric kernels is, under conditions discussed in Barndorff-Nielsen et al. (in press) and in Section 2 above, unbiased in a finite sample. The bias-corrected two-scale estimator is also unbiased under similar conditions. Not surprisingly, the required optimal (in a finite sample) number of autocovariances is, on average, smaller than in the unadjusted two-scale case (7, 6, and 10 vs. 16 in the case of the flat-top kernel estimators and 7 vs. 16 in the case of the bias-corrected two-scale estimator). Also, again not surprisingly, the difference between finite sample and asymptotic choices is smaller in the case of estimators with no obvious biases. If the conditions under which Bandi and Russell (2005), Zhang et al. (2005) and Barndorff-Nielsen et al. (in press) prove the finite sample and limiting properties of these estimators are satisfied at least as an approximation ($p^* \perp\!\!\!\perp \eta$ and i.i.d. η's for asymptotic results, $p^* \perp\!\!\!\perp \eta$, i.i.d. η's, and $W \perp\!\!\!\perp \sigma$ for finite sample results[14]), then (1) the bias-corrected two-scale estimator and the class of flat-top symmetric kernels should have no systematic biases, (2) the optimized (using finite sample Methods) bias-corrected two-scale estimator and flat-top symmetric kernels should have a smaller variance than their asymptotically optimal counterparts, (3) the optimized (using finite sample Methods) unadjusted two-scale estimator should have a smaller bias component than the asymptotically optimal unadjusted two-scale estimator.

The time variation in the optimal sampling intervals of the realized variance estimator and in the optimal (asymptotic and finite sample) number of autocovariances of the kernel estimators is substantial.[15] Since the optimal number of autocovariances is positively related to the number of trades in each day, for all kernel estimators we experience a clear upward trend in the required number of autocovariances reflecting the corresponding upward trend in the number of trades. Similarly, we experience a downward trend in the optimal sampling intervals of the realized variance estimator largely due to decreases in the noise second moment relative to the quarticity of the underlying equilibrium price process.

In the case of the unadjusted two-scale estimator, the optimal (in a finite sample) bandwidth choice is almost always higher than the asymptotically optimal choice. The only exceptions are for values at the end of the sample (for which, of course, asymptotic approximations are expected to fare better by virtue of the corresponding higher number of observations). As pointed out earlier, in light of the absence of a large finite sample bias component, asymptotic and finite sample choices are expected to be more similar in the case of the flat-top symmetric kernels and in the case of the bias-corrected two-scale estimator. This is what we find. In both cases, large deviations are generally associated with a lower, asymptotically optimal number of autocovariances.

The realized variance estimates tend to lead to higher forecasts than the flat-top symmetric kernel estimates and the bias-corrected two-scale estimates. Similarly, the asymptotically optimal unadjusted two-scale estimator lead to lower forecasts. This is indicative of a likely upward bias in the realized variance estimates and, again, a likely downward bias in the asymptotically optimal unadjusted two scale estimates. We now turn to option pricing and the profits from trading.

5. Profit-based ranking

We begin by considering options with a 6-hour expiration time. We rank the Methods based on average profits (Table 2) and Sharpe ratios (Table 3). The Sharpe ratios are simply defined as the average profits divided by the standard deviations of the profits. In what follows, the symbols $\overline{V}1000$, $\overline{V}1500$, and $\overline{V}1500M$, define the generic method $\overline{V}$ with ARMA parameters estimated using the three sample lengths described in Section 2, i.e., 1000 observations, 1500 observations, and the entire sample of data with at least 1500 observations, respectively.

When ranked based on average profits, 3 Methods perform better than Mean, namely $V_{BR}^{BNHLS(Bart)}1500$, $V_{BR}^{BNHLS(Cubic)}1500$, and

[14] As stressed above, the additional assumption used by Bandi and Russell (2005) in the case of the two-scale estimator, namely $\sigma_j^2 = \int_{t_{j-1}}^{t_j} \sigma_s^2 ds = \frac{V}{M} \forall j$, is simply intended to facilitate the empirical tractability of the resulting finite sample MSE expansion.

[15] We refer the interested reader to the unpublished, longer working paper version of this paper (Bandi et al., 2007) for graphical representations and additional details.

Table 2
Rank by annualized daily profits (in cents): 6-hour options

	Avg. prof.	HAC std.	Rank		Avg. prof.
OptRV1000	−0.1149	2.4576	30	Forecast length 1000	
OptRV1500	−0.6360	2.5044	33	FinFlatCubic1000	4.6568
OptRV1500M	−2.1653	2.6680	34	FinFlatBart1000	4.1716
Bart1000	2.5347	2.2001	17	FinFlatTukey1000	2.9372
Bart1500	3.6490	1.5335	9	FlatCubic1000	2.8112
Bart1500M	1.1831	1.9742	25	Bart1000	2.5347
FinZMA1000	2.2769	2.3014	19	FinZMA1000	2.2769
FinZMA1500	4.1112	1.4375	6	FlatBart1000	1.9480
FinZMA1500M	1.7384	1.8121	21	FlatTukey1000	1.7262
FlatBart1000	1.9480	2.3277	20	15minRV1000	0.7428
FlatBart1500	3.9282	1.5659	7	FinZMAadj1000	0.5479
FlatBart1500M	2.2990	1.4695	18	OptRV1000	−0.1149
FlatCubic1000	2.8112	1.6944	16	ZMAadj1000	−2.2520
FlatCubic1500	3.1022	1.4250	11	ZMA1000	−4.7069
FlatCubic1500M	0.9878	1.8519	26	5minRV1000	−5.1638
FlatTukey1000	1.7262	1.8137	22	Forecast length 1500	
FlatTukey1500	1.5302	1.6028	23	FinFlatBart1500	5.7373
FlatTukey1500M	−0.2343	1.8894	31	FinFlatCubic1500	5.0170
ZMA1000	−4.7069	3.6191	39	FinZMA1500	4.1112
ZMA1500	−3.2457	3.4138	37	FlatBart1500	3.9282
ZMA1500M	−2.6136	3.2316	36	Bart1500	3.6490
5minRV1000	−5.1638	2.8213	41	FlatCubic1500	3.1022
5minRV1500	−12.2026	3.4791	43	FinZMAadj1500	3.0726
5minRV1500M	−15.6579	3.7250	44	FinFlatTukey1500	2.9717
15minRV1000	0.7428	2.4936	27	FlatTukey1500	1.5302
15minRV1500	−3.4578	2.7453	38	ZMAadj1500	0.5084
15minRV1500M	−5.0107	2.8004	40	OptRV1500	−0.6360
FinFlatBart1000	4.1716	2.0844	5	ZMA1500	−3.2457
FinFlatBart1500	5.7373	1.1920	1	15minRV1500	−3.4578
FinFlatBart1500M	3.8587	1.2782	8	5minRV1500	−12.2026
FinFlatCubic1000	4.6568	1.6310	3	Forecast length >1500	
FinFlatCubic1500	5.0170	1.2324	2	FinFlatBart1500M	3.8587
FinFlatCubic1500M	3.0646	1.5726	13	FinZMAadj1500M	3.2501
FinFlatTukey1000	2.9372	1.6889	15	FinFlatCubic1500M	3.0646
FinFlatTukey1500	2.9717	1.2922	14	FlatBart1500M	2.2990
FinFlatTukey1500M	1.3474	1.5591	24	FinZMA1500M	1.7384
ZMAadj1000	−2.2520	2.8441	35	FinFlatTukey1500M	1.3474
ZMAadj1500	0.5084	2.5018	29	Bart1500M	1.1831
ZMAadj1500M	−0.4525	2.3944	32	FlatCubic1500M	0.9878
FinZMAadj1000	0.5479	2.5795	28	FlatTukey1500M	−0.2343
FinZMAadj1500	3.0726	1.9956	12	ZMAadj1500M	−0.4525
FinZMAadj1500M	3.2501	1.7353	10	OptRV1500M	−2.1653
Max	−16.8509	3.8173	45	ZMA1500M	−2.6136
Min	−5.3163	3.8745	42	15minRV1500M	−5.0107
Mean	4.4162	0.7528	4	5minRV1500M	−15.6579

Joint test all b's (profits) are equal to zero: $b' * \mathrm{inv}(\mathrm{var}(b)) * b = 249.2$
Pair-wise t-tests of equality between finite sample profits and asymptotic profits:

	1000	1500	>1500
t(FinFlatBart–FlatBart)	1.73	1.72	1.42
t(FinFlatCubic–FlatCubic)	1.75	2.82	2.80
t(FinFlatTukey–FlatTukey)	2.49	2.16	2.45
t(FinZMA–ZMA)	2.19	2.23	1.08
t(FinZMAadj–ZMAadj)	3.11	3.29	2.70
t(ZMAadj–ZMA)	1.90	3.01	1.64

$V_{BR}^{BNHLS(Cubic)}1000$. The cross-sectional dispersion of the profits is substantial, thereby indicating economically significant differences between alternative variance forecasts. Even though there are no systematic biases in the forecasts (Max and Min give profits equal to −16.85 and −5.3 with rankings equal to 45 and 42, respectively), upward biases are associated with the realized variance measures ($\widehat{V}^{Opt}$, $\widehat{V}^{5\,\min}$, and $\widehat{V}^{15\,\min}$), while downward biases are associated with $\widehat{V}_{ZMA}^{ZMA}$. The windows on the right of Table 2 break down the Methods based on their data lengths. In all cases the largest profits are given by the optimized (in a finite sample) flat-top symmetric kernels. The performance of the asymptotically optimal flat-top symmetric kernels is quite satisfactory, mainly in the Bartlett and cubic cases. The finite sample optimal (bias-corrected and unadjusted) two-scale estimators perform well and, sometimes, as well as the optimized (in a finite sample) flat-top symmetric kernels. Consistent with MSE-based arguments laid out in Section 2, when using an asymptotic bandwidth choice, bias-correcting the two-scale estimator is beneficial. For the longest data lengths, $\widehat{V}^{Opt}$ performs better than $\widehat{V}^{5\,\min}$ and $\widehat{V}^{15\,\min}$.

The standard deviation of the profits is quite variable across different Methods. The smallest standard deviation is associated with Mean, thus suggesting some degree of independence between alternative forecasts. Not surprisingly, ranking based on Sharpe ratios favors the Mean (see Table 3). The subsequent four Methods in the ranking are $\widehat{V}_{BR}^{BNHLS(Bar)}1500$, $\widehat{V}_{BR}^{BNHLS(Cub)}1500$, $\widehat{V}_{BR}^{BNHLS(Bar)}1500M$, and $\widehat{V}_{BR}^{ZMA}1500$. Importantly, when we break down the Methods on the basis of data length, we largely find the same ranking as earlier. The largest Sharpe ratios are generally associated with the optimized (in a finite sample) flat-top

Table 3
Rank by Sharpe ratios: 6-hour options

	Avg. prof./HAC std.	Rank		Avg. prof./HAC std.
OptRV1000	−0.0468	30	Forecast length 1000	
OptRV1500	−0.2540	33	FinFlatCubic1000	2.8552
OptRV1500M	−0.8116	36	FinFlatBart1000	2.0014
Bart1000	1.1521	18	FinFlatTukey1000	1.7391
Bart1500	2.3796	8	FlatCubic1000	1.6591
Bart1500M	0.5993	25	Bart1000	1.1521
FinZMA1000	0.9893	19	FinZMA1000	0.9893
FinZMA1500	2.8600	5	FlatTukey1000	0.9518
FinZMA1500M	0.9593	20	FlatBart1000	0.8369
FlatBart1000	0.8369	24	15minRV1000	0.2979
FlatBart1500	2.5085	7	FinZMAadj1000	0.2124
FlatBart1500M	1.5645	16	OptRV1000	−0.0468
FlatCubic1000	1.6591	15	ZMAadj1000	−0.7918
FlatCubic1500	2.1770	10	ZMA1000	−1.3006
FlatCubic1500M	0.5334	26	5minRV1000	−1.8303
FlatTukey1000	0.9518	22	Forecast length 1500	
FlatTukey1500	0.9547	21	FinFlatBart1500	4.8134
FlatTukey1500M	−0.1240	31	FinFlatCubic1500	4.0708
ZMA1000	−1.3006	39	FinZMA1500	2.8600
ZMA1500	−0.9508	37	FlatBart1500	2.5085
ZMA1500M	−0.8088	35	Bart1500	2.3796
5minRV1000	−1.8303	42	FinFlatTukey1500	2.2996
5minRV1500	−3.5074	43	FlatCubic1500	2.1770
5minRV1500M	−4.2035	44	FinZMAadj1500	1.5397
15minRV1000	0.2979	27	FlatTukey1500	0.9547
15minRV1500	−1.2595	38	ZMAadj1500	0.2032
15minRV1500M	−1.7893	41	OptRV1500	−0.2540
FinFlatBart1000	2.0014	11	ZMA1500	−0.9508
FinFlatBart1500	4.8134	2	15minRV1500	−1.2595
FinFlatBart1500M	3.0188	4	5minRV1500	−3.5074
FinFlatCubic1000	2.8552	6	Forecast length > 1500	
FinFlatCubic1500	4.0708	3	FinFlatBart1500M	3.0188
FinFlatCubic1500M	1.9488	12	FinFlatCubic1500M	1.9488
FinFlatTukey1000	1.7391	14	FinZMAadj1500M	1.8729
FinFlatTukey1500	2.2996	9	FlatBart1500M	1.5645
FinFlatTukey1500M	0.8642	23	FinZMA1500M	0.9593
ZMAadj1000	−0.7918	34	FinFlatTukey1500M	0.8642
ZMAadj1500	0.2032	29	Bart1500M	0.5993
ZMAadj1500M	−0.1890	32	FlatCubic1500M	0.5334
FinZMAadj1000	0.2124	28	FlatTukey1500M	−0.1240
FinZMAadj1500	1.5397	17	ZMAadj1500M	−0.1890
FinZMAadj1500M	1.8729	13	ZMA1500M	−0.8088
Max	−4.4144	45	OptRV1500M	−0.8116
Min	−1.3721	40	15minRV1500M	−1.7893
Mean	5.8663	1	5minRV1500M	−4.2035

symmetric kernels. The optimized (in a finite sample) two-scale estimators performs well. Again, when choosing an asymptotically optimal bandwidth, bias-correcting the two-scale estimator is helpful. As earlier, $\widehat{V}^{Opt}$ outperforms $\widehat{V}^{5\,\min}$ and $\widehat{V}^{15\,\min}$ for the longest data lengths.

We now deal with the overnights explicitly by multiplying each variance estimate by ζ and price 1-day options (Table 4).[16] Naturally, this adjustment has the potential to partly reduce the economic significance of alternative high-frequency volatility estimates/forecasts. Despite the adjustment, the overall picture remains fairly unchanged. Mean is now the 5th best forecast. The optimized (in a finite sample) flat-top symmetric kernels largely dominate all other forecasts for every choice of sample length. The asymptotically optimal flat-top symmetric kernels, the optimized (in a finite sample) "near consistent" Bartlett kernel, and the optimized (in a finite sample) two-scale estimators continue to fare well. Similarly, $\widehat{V}^{Opt}$ continues to outperform $\widehat{V}^{5\,\min}$ and $\widehat{V}^{15\,\min}$. Interestingly, the asymptotically optimal unadjusted two-scale estimator is now outperformed by all realized variance estimators. This is consistent with findings in Bandi and Russell (2005) where the same adjustment for lack of overnights is employed. Bias-correcting the asymptotically optimal two-scale estimator is again useful.[17] Examining the Sharpe ratios,[18] rather than the average profits, does not modify our results.

Importantly, the differences between alternative average profits/Sharpe ratios can be quite large. This result is consistent across experiments and indicates that, from an economic standpoint, there is scope for employing finite sample adjustments even in situations where the performance of the asymptotically optimal estimators is, in terms of ranking, similar to the performance of the optimal (in finite samples) estimators.

Not only are the profits economically significant, they are also statistically significant. Tables 2 and 4 contain the corresponding tests. Joint Chi-squared tests of the null of zero profits reject the null overwhelmingly both when considering 6-hour options and

[16] In Bandi et al. (2007) we also deal with the overnights by adding the square of the overnight returns to the variance estimates before forecasting. Our findings (below) are not affected by this alternative procedure.

[17] By construction, due to the adjustment, the unconditional Mean of the variance estimates is the same across Methods. Hence, it is the *time-variation* of the bias of the unadjusted two-scale estimator which makes the bias-correction compelling.

[18] The corresponding results can be obtained by the authors upon request.

Table 4
Rank by annualized daily profits (in cents): 1-day options

	Avg. prof.	HAC std.	Rank		Avg. prof.
OptRV1000	0.2282	2.8843	25	Forecast length 1000	
OptRV1500	1.4741	2.5052	19	FinFlatBart1000	4.9687
OptRV1500M	1.0151	2.4773	20	FinFlatCubic1000	4.9313
Bart1000	0.2227	2.6974	26	FinFlatTukey1000	2.9988
Bart1500	1.4768	2.2765	18	FlatCubic1000	2.8895
Bart1500M	1.6577	1.9858	14	FlatTukey1000	0.7762
FinZMA1000	0.0239	2.6797	28	OptRV1000	0.2282
FinZMA1500	1.5893	2.2472	15	Bart1000	0.2227
FinZMA1500M	1.4842	2.0402	17	FlatBart1000	0.1877
FlatBart1000	0.1877	2.1275	27	FinZMA1000	0.0239
FlatBart1500	−0.5808	1.9014	32	15minRV1000	−0.5118
FlatBart1500M	−1.1129	2.0240	33	5minRV1000	−1.6036
FlatCubic1000	2.8895	2.2533	11	FinZMAadj1000	−2.5947
FlatCubic1500	4.2241	1.5021	4	ZMAadj1000	−4.8756
FlatCubic1500M	3.0021	1.3820	9	ZMA1000	−7.6278
FlatTukey1000	0.7762	2.1234	23	Forecast length 1500	
FlatTukey1500	1.5229	1.6521	16	FinFlatCubic1500	4.5489
FlatTukey1500M	−0.2860	1.7227	30	FlatCubic1500	4.2241
ZMA1000	−7.6278	3.2403	43	FinFlatTukey1500	3.3046
ZMA1500	−7.4413	3.1052	42	FinFlatBart1500	3.2974
ZMA1500M	−7.7843	3.1828	44	FinZMA1500	1.5893
5minRV1000	−1.6036	3.2693	34	FlatTukey1500	1.5229
5minRV1500	0.8966	2.8442	21	Bart1500	1.4768
5minRV1500M	−0.1504	2.8772	29	OptRV1500	1.4741
15minRV1000	−0.5118	2.9695	31	5minRV1500	0.8966
15minRV1500	0.7834	2.8291	22	15minRV1500	0.7834
15minRV1500M	0.5232	2.6352	24	FlatBart1500	−0.5808
FinFlatBart1000	4.9687	1.8925	1	FinZMAadj1500	−1.8392
FinFlatBart1500	3.2974	1.2197	8	ZMAadj1500	−4.7779
FinFlatBart1500M	2.8223	1.4189	13	ZMA1500	−7.4413
FinFlatCubic1000	4.9313	2.1086	2	Forecast length >1500	
FinFlatCubic1500	4.5489	1.4320	3	FinFlatCubic1500M	3.7632
FinFlatCubic1500M	3.7632	1.3679	6	FlatCubic1500M	3.0021
FinFlatTukey1000	2.9988	2.1593	10	FinFlatTukey1500M	2.8380
FinFlatTukey1500	3.3046	1.3640	7	FinFlatBart1500M	2.8223
FinFlatTukey1500M	2.8380	1.3147	12	Bart1500M	1.6577
ZMAadj1000	−4.8756	2.7887	39	FinZMA1500M	1.4842
ZMAadj1500	−4.7779	2.5778	38	OptRV1500M	1.0151
ZMAadj1500M	−5.0953	2.7134	40	15minRV1500M	0.5232
FinZMAadj1000	−2.5947	2.3755	37	5minRV1500M	−0.1504
FinZMAadj1500	−1.8392	2.2005	35	FlatTukey1500M	−0.2860
FinZMAadj1500M	−2.3207	2.3607	36	FlatBart1500M	−1.1129
Max	−7.9026	3.8924	45	FinZMAadj1500M	−2.3207
Min	−6.4832	3.9416	41	ZMAadj1500M	−5.0953
Mean	3.9504	0.8150	5	ZMA1500M	−7.7843

Joint test all b's (profits) are equal to zero: $b' * \text{inv}(\text{var}(b)) * b = 85.5$
Pair-wise t-tests of equality between finite sample profits and asymptotic profits:

	1000	1500	>1500
t(FinFlatBart–FlatBart)	2.52	2.51	2.31
t(FinFlatCubic-FlatCubic)	1.79	0.35	0.85
t(FinFlatTukey-FlatTukey)	2.44	2.25	2.56
t(FinZMA-ZMA)	1.79	2.35	2.36
t(FinZMAadj-ZMAadj)	1.98	2.74	2.77
t(ZMAadj-ZMA)	2.10	3.31	3.38

when considering 1-day options. The 5% critical value of a Chi-squared test with 45 degrees of freedom, as in our case, is about 61. The tests' values are 249.2 and 85.5, respectively. Hence, at least one strategy earns profits statistically significantly different from zero. Pairwise t-tests of the null of equal profits between optimal (in a finite sample) kernel estimators and asymptotically optimal kernel estimators favor the former unequivocally. The profits are *always* larger in the case of the optimal (in a finite sample) kernel estimators and generally statistically so. Interestingly, even in situations where a certain asymptotically optimal kernel estimator performs as well as (or better) than an alternative optimal (in a finite sample) kernel estimator, the corresponding optimal (in a finite sample) version of that estimator delivers profits that are, in general, statistically higher. Hence, it is valuable to optimize the performance of estimators that would fare satisfactorily even when implemented using asymptotic bandwidth selection criteria. Finally, the asymptotically optimal bias-corrected two-scale estimator always outperforms its unadjusted asymptotically optimal counterpart while yielding gains that are generally statistically significant.[19] To summarize, we find that:

[19] Although the goal of this paper is to compare alternative high-frequency variance forecasts, it is tempting to evaluate how these forecasts would fare against a readily-available market-based forecast, such as VIX, in the context of our metrics. At the suggestion of one of the referees (whom we thank), in Bandi et al. (2007) we use VIX as a volatility forecast and associate it with an additional trader (after an appropriate adjustment for potential risk premia in volatility). We find that the resulting ranking is hardly affected by the inclusion of VIX. In addition, for the case of daily options, VIX ranks 38th out of 46 Methods.

1. Optimized (in a finite sample) flat-top symmetric kernels generally outperform other choices.
2. Within the class of flat-top symmetric kernels, the Bartlett and cubic kernel are the preferred choices in most of our experiments. The modified Tukey-Hanning kernel is virtually always inferior to these alternatives, despite its favorable theoretical properties.
3. The asymptotically optimal flat-top symmetric kernels, the optimized (in a finite sample) "near consistent" Bartlett kernel, and the finite sample optimal (adjusted or unadjusted) two-scale estimators can perform very well.
4. When using asymptotically optimal bandwidth choices, bias-correcting the two-scale estimator is always beneficial for our data and metrics.
5. Importantly, *regardless of the estimator*, optimal (in a finite sample) bandwidth choices *always* yield higher profits than the corresponding asymptotically optimal choices for our data. These higher profits are generally statistically significant.
6. Optimally-sampled realized variance generally dominates realized variance based on ad-hoc intervals.

We find it interesting, and of course not obvious, that our reported rankings almost perfectly mirror those found by Bandi and Russell (2005) using finite sample MSE expansions as the relevant metric. Bandi and Russell (2005) emphasize that evaluating volatility estimators solely based on their limiting properties is, in general, not the right criterion. Certain "near consistent" and consistent estimators, such as the Bartlett kernel estimator and the unadjusted two-scale estimator, have favorable asymptotic properties but can display large finite sample biases. Equivalently, estimators with the same asymptotic properties, such as the unadjusted two-scale estimator and the unbiased flat-top Bartlett kernel estimator, can have drastically different finite sample features, largely due to bias considerations. For a variety of relevant metrics, the (time-varying) finite sample bias needs to be reduced for certain estimators to perform at their full potential. Our option pricing metrics, for example, penalize forecasts which drastically overstate or understate the level of volatility. Finite sample bandwidth selection Methods, such as the one proposed by Bandi and Russell (2005), can yield bias reduction while optimally trading-off bias and variance. Because some consistent or "near consistent" estimators are asymptotically unbiased, while being biased in finite samples, asymptotic bandwidth selection Methods can be detrimental in practice. In general, these Methods do not capture the finite sample bias of these estimators. We also show that, even in the case of estimators with no obvious biases, such as the bias-corrected two-scale estimator and the class of flat-top kernel estimators, there is scope for optimizing their finite sample variance properties by appropriately selecting the corresponding number of autocovariances.

6. Conclusions

The relative success of alternative high-frequency variance measures depends on their out-of-sample forecasting performance. Yet, a variety of loss functions can be proposed in practice. Either statistical or economic metrics could be entertained. Within each broad category, several alternative criteria can be invoked. It is legitimate to argue that well-posed economic loss functions represent crucial measures of the success of different forecasts. We take this position in this paper. Specifically, we attempt to provide a rich comparison between recently-proposed high-frequency measures of variance in the context of important, nonlinear economic criteria.

In the context of our proposed metrics, we do not find a significant difference between the "ex-ante" ranking that one would derive from careful evaluation (and optimization) of the finite sample MSE properties of alternative estimators and the "ex-post" ranking of the estimators' out-of-sample forecasts.

Among other things, future work should evaluate the pricing and trading of longer-maturity options, and the resulting ranking of long-term volatility estimates/forecasts. It is also of interest to examine situations where pricing and trading decisions are affected by forecast precision, rather than solely by point forecasts. More importantly for our purposes, while we empirically show that MSE-based forecasts largely preserve their "ex-ante" ranking in the context of our metrics, the optimal forecasts are in general not the forecasts derived from MSE-optimal estimates. Direct optimization of certain economic criteria could then be very beneficial and is left as a crucial topic for future work.

References

Aït-Sahalia, Y., Mancini, L., 2008. Out-of-sample forecasts of quadratic variation. Journal of Econometrics 147 (1), 17–33.

Andersen, T.G., Bollerslev, T., Meddahi, N., 2006. Realized volatility forecasting and market microstructure noise. Working Paper.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Ebens, H., 2001. The distribution of realized stock return volatility. Journal of Financial Economics 61, 43–76.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 1999. (Understanding, optimizing, using, and forecasting) Realized volatility and correlation. Working Paper.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2000. Great realizations. Risk Magazine 13, 105–108.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71, 579–625.

Bandi, F.M., Russell, J.R., 2003. Microstructure noise, realized volatility, and optimal sampling. Working Paper.

Bandi, F.M., Russell, J.R., 2005. Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations. Working Paper.

Bandi, F.M., Russell, J.R., 2006. Separating microstructure noise from volatility. Journal of Financial Economics 79, 655–692.

Bandi, F.M., Russell, J.R., 2008a. Volatility. In: Birge, J., Linetski, V. (Eds.), Handbook of Financial Engineering. Elsevier, North Holland.

Bandi, F.M., Russell, J.R., 2008b. Microstructure noise, realized variance, and optimal sampling. Review of Economic Studies 75, 339–369.

Bandi, F.M., Russell, J.R., Yang, C., 2007. Realized volatility forecasting and option pricing. Technical Report. http://faculty.chicagogsb.edu/federico.bandi/research/BRY_forecasting_optionpricing.pdf.

Bandi, F.M., Russell, J.R., Zhu, Y., 2008. Using high-frequency data in dynamic portfolio choice. Econometric Reviews 27, 163–198.

Barndorff-Nielsen, O.E., Shephard, N., 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.

Barndorff-Nielsen, O.E., Shephard, N., 2007. Variation, jumps, market frictions and high-frequency data in financial econometrics. In: Blundell, R., Torsten, P., Newey, W.K. (Eds.), Advances in Economics and Econometrics. Theory and Applications. Ninth World Congress. Econometric Society Monographs. Cambridge University Press, pp. 328–372.

Barndorff-Nielsen, O.E., Hansen, P., Lunde, A., Shephard, N., 2005. Regular and modified kernel-based estimators of integrated variance: The case with independent noise. Working Paper.

Barndorff-Nielsen, O.E., Hansen, P., Lunde, A., Shephard, N., 2006. Designing realized kernels to measure ex-post variation of equity prices in the presence of noise. Econometrica (in press).

Barndorff-Nielsen, O.E., Hansen, P., Lunde, A., Shephard, N., 2007. Subsampling realized kernels. Working Paper.

Corradi, V., Distaso, W., Swanson, N., 2006. Predictive inference for integrated volatility. Working Paper.

Corsi,, 2003. A simple long-memory model of realized volatility. Working Paper.

De Pooter, M., Martens, M., Van Dijk, D., 2008. Predicting the daily covariance matrix for S&P100 stocks using intraday data: But which frequency to use? Econometric Reviews 27, 199–229.

Elton, E.J., Gruber, M., Comer, G., Li, K., 2002. Spiders: Where are the bugs? Journal of Business 75, 453–472.

Engle, R.F., Hong, C.H., Kane, A., 1990. Valuation of variance forecasts with simulated option markets. NBER Working Paper No: 3350.

Fleming, J., Kirby, C., Ostdiek, B., 2001. The economic value of volatility timing. Journal of Finance 56, 329–352.

Fleming, J., Kirby, C., Ostdiek, B., 2003. The economic value of volatility timing using "realized volatility". Journal of Financial Economics 67, 473–509.

Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. Journal of Time Series Analysis 4, 221–238.

Ghysels, E., Sinko, A., 2006a. Comment on Hansen and Lunde. Journal of Business and Economic Statistics 24, 192–194.

Ghysels, E., Sinko, A., 2006b. Volatility forecasting and microstructure noise. Working Paper.

Ghysels, E., Santa-Clara, P., Valkanov, R., 2006. Predicting volatility: Getting the most out of return data sampled at different frequencies. Journal of Econometrics 131, 59–95.
Hansen, P.R., Lunde, A., 2005. A realized variance for the whole day based on intermittent high-frequency data. Journal of Financial Econometrics 4, 525–554.
Hansen, P.R., Lunde, A., 2006. Realized variance and market microstructure noise. Journal of Business and Economic Statistics 24, 127–161 (with discussions).
Kalnina, I., Linton, O., 2008. Estimating quadratic variation consistently in the presence of correlated measurement error. Journal of Econometrics 147 (1), 47–59.
Large,, 2006. Estimating quadratic variation when quoted prices change by constant increments. Working Paper.
McAleer, M., Medeiros, M., 2008. Realized volatility: A review. Econometric Reviews 27, 10–45.
Oomen, R.C.A., 2005. Properties of bias-corrected realized variance under alternative sampling schemes. Journal of Financial Econometrics 3, 555–577.
Protter, P., 1995. Stochastic Integration and Differential Equations. A New Approach. Springer.
West, K.D., Edison, H.J., Cho, D., 1993. A utility-based comparison of some models of exchange rate volatility. Journal of International Economics 35, 23–45.
Zhang, L., Mykland, P., Aït-Sahalia, Y., 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. Journal of the American Statistical Association 100, 1394–1411.
Zhang, L., 2006. Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. Bernoulli 12, 1019–1043.
Zhou, B., 1996. High-frequency data and volatility in foreign-exchange rates. Journal of Business and Economic Statistics 14, 45–52.