Realized volatility forecasting in the presence of time-varying noise*

Federico M. Bandi,*** Jeffrey R. Russell,** Chen Yang*
***Johns Hopkins University and Edhec-Risk,**University of Chicago, *Moody's

May 2010

Abstract

Observed high-frequency financial prices can be considered as comprising two components, a true price and a market microstructure noise perturbation. It is an empirical fact that the second moment of market microstructure noise is time-varying. We study the optimal design of nonparametric variance estimators in linear variance forecasting models with time-varying market microstructure noise. Specifically, we discuss optimal frequency selection in the case of the classical realized variance estimator and optimal bandwidth selection in the case of kernel-type integrated variance estimators. We show that the sampling frequencies (bandwidths) are generally considerably lower (larger) than those that would be optimally chosen in linear forecasting models when time-variation in the second moment of the noise is unaccounted for. In this setting, we suggest treating the relevant frequency/bandwidth as a parameter and estimating it *jointly* with the parameters of the forecasting model. Conditional frequency/bandwidth choices are also discussed. Conditions guaranteeing their superior performance over the unconditional choices are derived.

Keywords: Realized Variance, Kernel-based Estimators, Time-Varying Market Microstructure Noise, Volatility Forecasting.

*We thank Torben Andersen, Nour Meddahi, and seminar participants at the Imperial College Workshop on High-Frequency Data (London, February 22, 2007), the Conference on Volatility and High-Frequency Data (Chicago, April 21-22, 2007), and the 75th Cowles Foundation Anniversary Conference "Looking to the Future: A New Generation of Econometricians" (New Haven, June 11-12, 2007) for discussions. We are grateful to the William S. Fishman Faculty Research Fund at Chicago Booth, University of Chicago, and Carey Business School, Johns Hopkins University (Bandi), NYU Stern (Russell), and Chicago Booth, University of Chicago (Bandi, Russell, and Yang) for financial support. Correspondence can be sent to: fbandi1@jhu.edu, jeffrey.russell@chicagobooth.edu, michelle.chen.yang@gmail.com.

1 Introduction

Observed high-frequency financial prices can be considered as comprising two components, a true price and a market microstructure noise perturbation. It is an empirical fact (consistent with market microstructure theory) that the second moment of market microstructure noise is time-varying (for evidence, Bandi and Russell, 2006a, Hansen and Lunde, 2006, and Oomen, 2006). This time variation (as illustrated in Figure 1) induces time variation in the bias of the realized variance estimator constructed using high-frequency data. Naturally, a time-varying bias in realized variance has implications for variance forecasting. This paper shows that the time-varying nature of market microstructure noise is a fundamental aspect of the variance forecasting problem in the presence of high-frequency financial prices. This is true both from a theoretical and an empirical perspective.

Optimal frequency selection (in the case of realized variance) and bandwidth selection (in the case of alternative kernel-type variance estimators) have been very active areas of recent research (Bandi and Russell, 2007, Barndorff-Nielsen and Shephard, 2007, and McAleer and Medeiro, 2008, are recent surveys on the subject). Joint consideration of the frequency/bandwidth selection and forecasting problem is pursued by Andersen et al. (2010), ABM hereafter, and Ghysels and Sinko (2010), GS henceforth. Specifically, in the context of theoretical linear regression models with time-invariant noise, they show that the optimal frequency/bandwidth selection problem (for the purpose of R^2 maximization) reduces (under assumptions) to the minimization of the unconditional variance of the estimator (regressor).

We derive novel optimal frequency/bandwidth selection methods for the joint frequency/bandwidth selection and forecasting problem in the presence of time-varying microstructure noise variance. In this context, we find considerably lower optimal frequencies and, similarly, considerably larger bandwidths than those obtained when time variation in the noise variance is unaccounted for. Interestingly, the new frequency choices are in general closer to those that would be derived from optimization of the finite sample unconditional mean-squared-error (MSE) of the regressor (as in Bandi and Russell, 2006a, 2008) than to those that would be obtained from optimization of the unconditional variance of the regressor (under the assumption of a constant noise second moment), as needed in linear forecasting models (ABM, 2010, and GS, 2010). Specifically, we find that taking bias into account sub-optimally through an unconditional MSE-based optimization (under the assumption of a constant noise second moment) can be an empirically reasonable strategy.

We however propose the superior strategy of jointly estimating the optimal (from an R^2 standpoint) frequency/bandwidth by least-squares along with the parameters of the forecasting model. This solution coincides with the R^2-optimal, closed-form, solution which we obtain in the context of our illustrative model (in Section 1 below). However, as we discuss in what follows, it applies to general forecasting models making use of realized variance measures and, importantly, is robust to features of the data such as noise dependence, dependence between the noise process and the price process, and dependence between the equilibrium price variance and the noise variance.

Finally, we study cases in which choosing the sampling frequency/bandwidth *conditionally* (for each entry/day in the regressor vector) rather than *unconditionally,* as often done in the literature, is a better strategy from a forecasting standpoint, even when the alternative is the exact, unconditional, R^2-based optimal choice. An application to SPIDERS mid-quotes validates our theoretical predictions.

Several promising, recent contributions have studied variance forecasting using microstructure noise-contaminated high-frequency variance estimates. These contributions have evaluated the forecasting potential of the classical realized variance estimator (Andersen et al., 2003, and Barndorff-Nielsen et al., 2002) as well as that of classes of theoretically more robust (to noise) kernel-based estimators (e.g., Zhou, 1996, Hansen and Lunde, 2006, Zhang et al., 2005, and Barndorff-Nielsen et al., 2008). Given time-series of alternative variance estimates, the forecasts have generally been obtained using ARFIMA models (Bandi and Russell, 2008, 2010, and Bandi et al., 2008), Mincer-Zarnowitz-style linear regressions (Andersen et al., 2010), or MIDAS-type regressions (Ghysels and Synko, 2006, 2010). Either statistical metrics, such as forecast MSEs and coefficients of determination (Aït-Sahalia and Mancini, 2008, Andersen et al., 2010, Corradi et al., 2006, and Ghysels and Sinko, 2006, 2010) or economic metrics, such as the utility obtained by investors or the profits obtained by option traders on the basis of alternative variance forecasts (Bandi and Russell, 2008, 2010, and Bandi et al., 2008), have been used for the purpose of evaluating the goodness of the forecasts.

Time-varying noise variance and its impact on variance forecasting have been considered by Bandi and Russell (2008, 2010) and Bandi et al. (2008) using *conditional* frequency/bandwidth selection rules. The joint frequency/bandwidth selection and forecasting problem has been studied by ABM (2010) and GS (2010) using *unconditional* optimal rules in the presence of a time-invariant noise second moment. In this paper, we emphasize the importance of the time-varying nature of the noise variance for the joint problem, we provide an appealing least-squares solution to the unconditional problem, and discuss conditions under which conditional optimality methods may be preferable to unconditional methods.

We work with a price formation mechanism, presented in Section 1, which has been broadly adopted in the literature and is extended here solely to allow for a time-varying noise second moment. Similarly, we illustrate matters in the context of a simple, but commonly-employed, autoregressive forecasting model. Both choices are meant to derive closed-form results and illustrate important issues pertaining to variance forecasting using realized variance measures when the noise variance is time-varying. In the unconditional case, the theoretical analysis will however lead us to a least-squares solution to the choice problem which, as emphasized in Section 9, has general applicability and does not hinge either on the fine-grain features of the relation between noise and price process or on the forecasting model.

In what follows, we use the symbol $\perp\!\!\!\perp$ to signify "statistical independence."

2 A classical price formation mechanism and a forecasting model

Consider a trading day t. Assume availability of $M+1$ equispaced, observed logarithmic asset prices over $[t, t+1]$ and write

$$p_{t+j\delta} \;=\; p^*_{t+j\delta} + u_{t+j\delta} \qquad j = 0, ..., M$$

or, in terms of continuously-compounded returns,

$$\underbrace{p_{t+j\delta} - p_{t+(j-1)\delta}}_{r_{t+j\delta}} \;=\; \underbrace{p^*_{t+j\delta} - p^*_{t+(j-1)\delta}}_{r^*_{t+j\delta}} + \underbrace{u_{t+j\delta} - u_{t+(j-1)\delta}}_{\varepsilon_{t+j\delta}}, \qquad j = 1, ..., M,$$

where p^* denotes the *unobservable* true price, u denotes *unobservable* market microstructure noise, and $\delta = \frac{1}{M}$ represents the time distance between adjacent price observations.

We assume the true price process evolves in time as a stochastic volatility local martingale, i.e.,

$$p^*_t = \int_0^t \sigma_s dW_s,$$

where σ_t is a càdlàg stochastic volatility process, W_t is a standard Brownian motion, and $\sigma \perp\!\!\!\perp W$ (leverage effects are ruled out[1]). The daily integrated variance is therefore defined as $V_{t,t+1} = \int_t^{t+1} \sigma_s^2 ds$. From now on, we abuse notation a bit and write V_t instead of $V_{t,t+1}$.

We assume the noise contaminations in the price process u are IID in discrete time, over each day, with variance σ^2_{tu} and fourth moment $c_u \sigma^4_{tu}$ (where c_u denotes kurtosis). In addition, $u \perp\!\!\!\perp p^*$ and $\sigma^2_{tu} \perp\!\!\!\perp V_t$ (the latter assumption will be relaxed in Section 7).

Importantly, the variance of the market microstructure noise has a subscript t to signify that it can change from day to day. All other assumptions are classical assumptions in this literature and, with the exception of $\sigma \perp\!\!\!\perp W$ (which can be easily relaxed in asymptotic designs), have been routinely employed when studying the finite sample and asymptotic properties of nonparametric estimates of integrated variance in the presence of noise (see, e.g., ABM (2010), Bandi and Russell, 2003, 2010, Barndorff-Nielsen et al., 2008, Hansen and Lunde, 2006, and Zhang et al., 2005, among others).[2] While the assumptions capture important first-order effects in the data, their empirical accuracy depends on the market structure, on the price measurement (transaction prices vs. mid-quotes, for example), as well as on the sampling scheme (calendar time vs. event time, for instance). Bandi and Russell (2006b) discuss these ideas. Here, we depart from the "usual" assumptions by allowing for a time-varying noise second moment. Assuming a constant second moment over a day, but allowing it to change from day to day, is a useful way to combine theoretical soundness with

[1] Bandi and Russell (2006b) discuss the validity of this assumption in the case of equities and exchange rates.

[2] Aït-Sahalia et al. (2005), Bandi and Russell (2003, 2008), Hansen and Lunde (2006), and Oomen (2005, 2006) discuss noise dependence. Kalnina and Linton (2008) allow for a form of dependence between the noise and the true price.

empirical tractability. As we discuss below, the time-varying second moments can be estimated consistently (nonparametrically) for each day in the sample. Alternatively, one could imagine a situation where *each* noise contamination is endowed with a time-varying second moment.[3] We leave the theoretical and empirical complications that this modelling choice would entail for future work.

We are interested in predicting V_{t+1} given past daily values of the classical realized variance estimator, namely $\widehat{V}_t = \sum_{j=1}^{M} r_{t+j\delta}^2$ (Andersen et al., 2003, and Barndorff-Nielsen et al., 2002). To this extent, it is useful to begin with a specific model which forms the basis for some of our analysis in the paper, although general results will be presented later. Assume $V_{t+1} = \alpha + \beta V_t + \xi_{t+1}$, where ξ_{t+1} is such that $\mathbf{E}(\xi_{t+1}|\mathfrak{F}_t) = 0$. The model estimation is performed using lagged values of $\widehat{V}_t$, leading to the forecasting regression

$$V_{t+1} = \widehat{\alpha} + \widehat{\beta}\widehat{V}_t + \widehat{\xi}_{t+1}. \tag{1}$$

The next section provides intuition about the main effects of time-varying noise on the sampling frequency and the model's parameters.

3 Preliminary discussion and intuition

Under our assumed structure, the realized variance estimator takes the form $\widehat{V}_t = \sum_{j=1}^{M} r_{t+j\delta}^2 = \sum_{j=1}^{M} r_{t+j\delta}^{*2} + \sum_{j=1}^{M} \varepsilon_{t+j\delta}^2 + \sum_{j=1}^{M} r_{t+j\delta}^{*}\varepsilon_{t+j\delta}$. The estimator can also be rewritten as $\widehat{V}_t = V_t + M\mathbf{E}\left(\sigma_{t\varepsilon}^2\right) + \left(\sum_{j=1}^{M} r_{t+j\delta}^{*2} - V_t\right) + M\left(\sigma_{t\varepsilon}^2 - \mathbf{E}\left(\sigma_{t\varepsilon}^2\right)\right) + \left(\sum_{j=1}^{M} \varepsilon_{t+j\delta}^2 - M\sigma_{t\varepsilon}^2\right) + \sum_{j=1}^{M} r_{t+j\delta}^{*}\varepsilon_{t+j\delta}$. We denote the estimation error with no market microstructure noise by $a_t = \sum_{j=1}^{M} r_{t+j\delta}^{*2} - V_t$. We define the difference between the realized bias on day t and the expected bias on day t as $\widetilde{a}_t = \left(\sum_{j=1}^{M} \varepsilon_{t+j\delta}^2 - M\sigma_{t\varepsilon}^2\right)$. The unconditional expected bias is given by $M\mathbf{E}\left(\sigma_{t\varepsilon}^2\right)$ and the difference between the expected day-t bias and the unconditional bias is given by $M\left(\sigma_{t\varepsilon}^2 - \mathbf{E}\left(\sigma_{t\varepsilon}^2\right)\right)$. Finally, we denote the mean-zero cross-product term by $\gamma_t = \sum_{j=1}^{M} r_{t+j\delta}^{*}\varepsilon_{t+j\delta}$.

The forecast error of the estimated model can now be expressed as $\widehat{\xi}_{t+1} = V_{t+1} - \left(\widehat{\alpha} + \widehat{\beta}\widehat{V}_t\right) = \alpha + \beta V_t + \xi_{t+1} - \widehat{\alpha} - \widehat{\beta}\left[V_t + M\mathbf{E}\left(\sigma_{t\varepsilon}^2\right) + a_t + M\left(\sigma_{t\varepsilon}^2 - \mathbf{E}\left(\sigma_{t\varepsilon}^2\right)\right) + \widetilde{a}_t + \gamma_t\right]$. Re-arranging terms yields $\widehat{\xi}_{t+1} = \left[\alpha - \widehat{\alpha} - \widehat{\beta}M\mathbf{E}\left(\sigma_{t\varepsilon}^2\right)\right] + \left(\beta - \widehat{\beta}\right)V_t - \widehat{\beta}\left(a_t + M\left(\sigma_{t\varepsilon}^2 - \mathbf{E}\left(\sigma_{t\varepsilon}^2\right)\right) + \widetilde{a}_t + \gamma_t\right) + \xi_{t+1}$. For any values of M, $\widehat{\alpha}$,

[3] In this case, our estimates may be readily interpreted as local daily averages. Importantly, our estimates can be further localized in the sense that, at the cost of decreased accuracy, the noise second moment can be estimated consistently (under our assumptions) over any fixed intra-daily period.

and $\widehat{\beta}$, the variance of the forecast error is now given by

$$\mathbf{Var}\left(\widehat{\xi}_{t+1}\right) = \left(\beta-\widehat{\beta}\right)^2\mathbf{Var}\left(V_t\right)+\widehat{\beta}^2\left(\underbrace{\underbrace{\underbrace{\mathbf{Var}\left(a_t\right)}_{No\ noise}+\mathbf{Var}\left(\widetilde{a}_t\right)+\mathbf{Var}\left(\gamma_t\right)}_{Noise}+M^2\mathbf{Var}\left(\sigma^2_{t\varepsilon}\right)}_{Time-varying\ noise}\right)+\mathbf{Var}(\xi_{t+1}). \tag{2}$$

The sampling frequency only affects the second term in Eq. (2) implying that the value of M which minimizes the forecast error variance can be determined without consideration of the parameters of the forecasting model. Focusing on this second term, we note that the no noise case coincides with $\mathbf{Var}\left(a_t\right)$. This variance is minimized by choosing M as large as possible. The time-invariant noise case leads to $\mathbf{Var}\left(a_t\right)+\mathbf{Var}\left(\widetilde{a}_t\right)+\mathbf{Var}\left(\gamma_t\right)$, a concave function of the sampling frequency. The resulting optimal M is lower than the optimal M in the no noise case but, as we show formally in the next section, larger than the optimal M in the time-varying noise case due to the presence, in the latter case, of the extra term $M^2\mathbf{Var}\left(\sigma^2_{t\varepsilon}\right)$. In sum, the frequency which minimizes the forecast error variance coincides with the value which minimizes the variance of the regressor $(\widehat{V}_t)$. Below, we provide a representation for this variance in terms of the structural parameters.

The value of $\widehat{\beta}$ which minimizes the forecast error variance, instead, depends on the choice of M and, hence, on the variance of $\widehat{V}_t$. Importantly, if V_t were observable, we would have $\mathbf{Var}\left(\widehat{\xi}_{t+1}\right) = \left(\beta-\widehat{\beta}\right)^2\mathbf{Var}\left(V_t\right)+\mathbf{Var}(\xi_{t+1})$ and the solution to the problem would be classical: $\widehat{\beta}=\beta=\mathbf{Cov}(V_{t+1},V_t)/\mathbf{Var}(V_t)$. In our case,

$$\widehat{\beta}=\beta\frac{\mathbf{Var}(V_t)}{\mathbf{Var}(V_t)+\left(\mathbf{Var}\left(a_t\right)+\mathbf{Var}\left(\widetilde{a}_t\right)+M^2\mathbf{Var}\left(\sigma^2_{t\varepsilon}\right)+\mathbf{Var}\left(\gamma_t\right)\right)} \tag{3}$$

and the theoretical least-squares $\widehat{\beta}$ estimate is attenuated by the *optimized* (over M) variance of the mean-zero measurement error component in realized variance, thereby giving $\widehat{\beta}<\beta$.

Our results are presented in two parts. First, we consider methods for optimally-choosing M. In our set-up, we obtain (near) closed-form solutions for M. Next, we consider the joint problem of estimating the forecasting parameters along with M. Again, while closed-form solutions are available for the joint problem in the simple model presented here, the class of forecasting models, as well as the types of estimators of realized variance, is extended beyond our current set-up (in Section 9). Convergence rates of the estimated parameters are also established.

4 Optimal forecasting frequencies: closed-form expressions

Theorem 1 presents the optimal rule to choose the R^2-maximizing number of observations M.

Theorem 1.

Consider the regression in Eq. (1). *Then,*

$$\begin{aligned} M_1 &= \arg\max R_M^2 \\ &= \arg\min \left\{ \frac{2}{M}\mathbf{E}(Q_t) + (2\mathbf{E}(\theta_{t\varepsilon}) - 3\mathbf{E}(\sigma_{t\varepsilon}^4))M + M^2\mathbf{Var}(\sigma_{t\varepsilon}^2) \right\}, \end{aligned} \tag{4}$$

where $Q_t = \int_t^{t+1} \sigma_s^4 ds$, $\theta_{t\varepsilon} = \mathbf{E}(\varepsilon_t^4)$, *and* $\sigma_{t\varepsilon}^2 = \mathbf{E}(\varepsilon_t^2) = 2\sigma_{tu}^2$.

Proof.

See Appendix.

Remark 1. (Interpretation.)

M_1 minimizes the unconditional variance of the regressor (realized variance). Under an assumption of independence between σ_{tu}^2 and V_t (relaxed in Section 7), this minimization translates into maximization of the forecasting regression's R_M^2 (as in ABM, 2010, and GS, 2010). The form of this unconditional variance is unusual and includes a term (of order M^2) which accounts for the variability of the noise variance (i.e., the last term in Eq. (4)).

Remark 2. (Implementation.)

The quantities $\theta_{t\varepsilon}$ and $\sigma_{t\varepsilon}^2$ can be estimated consistently (for each day in the sample) by using sample moments of the observed return data sampled at the highest frequencies (Bandi and Russell, 2003, 2006a).[4] Given $\widehat{\theta}_{t\varepsilon}$ and $\widehat{\sigma}_{t\varepsilon}^2$, consistent estimates of the unconditional moments $\mathbf{E}(\theta_{t\varepsilon})$, $\mathbf{E}(\sigma_{t\varepsilon}^4)$, and $\mathbf{Var}(\sigma_{t\varepsilon}^2)$ can be obtained by employing sample moments of the daily estimates under classical (stationarity) assumptions. Estimation of the daily quarticity Q_t can be conducted by sampling the observed returns at relatively low (15- or 20-minute) frequencies.[5] Roughly unbiased estimates of the unconditional moment $\mathbf{E}(Q_t)$ can then be derived by averaging the estimated daily quarticities under an assumption of stationarity for Q_t. While empirical implementation of the method (by virtue of numerical minimization of the function in Eq. (4)) is fairly straightforward, the Corollary below provides a convenient, approximate rule to select the optimal M.

Corollary to Theorem 1.

Consider the regression in Eq. (1). *For a large optimal* M_1,

[4] Bandi and Russell (2007) discuss finite sample bias corrections.

[5] Bandi and Russell (2008) discuss the empirical validity of this simple (albeit theoretically inefficient) procedure by simulation. Efficient estimation of the quarticity is an important issue for future work.

$$M_1^* \approx \arg\max R_M^2 = \left(\frac{\mathbf{E}(Q_t)}{\mathbf{Var}(\sigma_{t\varepsilon}^2)}\right)^{1/3}. \tag{5}$$

Remark 3

The approximate rule in Eq. (5) readily adapts to the noise variance's variance. The larger this variance relative to the signal $\mathbf{E}(Q_t)$ coming from the underlying equilibrium price, the smaller the optimal number of observations needed to compute $\widehat{V}$. As always in these problems, a smaller number of observations translates into smaller noise contaminations.

Remark 4

The derived approximate rule differs from the optimal (in an unconditional finite sample MSE sense), approximate rule proposed by Bandi and Russell (2003, 2008) in the presence of time-invariant noise, i.e.,

$$M_2^* = \left(\frac{\mathbf{E}(Q_t)}{(\mathbf{E}(\varepsilon^2))^2}\right)^{1/3}. \tag{6}$$

It also differs from the optimal (in an R^2 sense), approximate rule proposed by ABM (2010) and GS (2010) in the case of time-invariant noise, i.e.,[6]

$$M_3^* = \left(\frac{2\mathbf{E}(Q_t)}{2\mathbf{E}(\varepsilon^4) - 3\mathbf{E}(\varepsilon^2)^2}\right)^{1/2}. \tag{7}$$

Naturally, the relative performance of these alternative rules depends on their relation with M_1^*. In general, $M_3^* > M_2^*$. This is easy to see. Since $2\mathbf{E}(\varepsilon^4) - 3\mathbf{E}(\varepsilon^2)^2 = 4\mathbf{E}(u^4)$, then $M_2^* = \left(\frac{\mathbf{E}(Q_t)}{4\sigma_u^4}\right)^{1/3}$ and $M_3^* = \left(\frac{2\mathbf{E}(Q_t)}{4c_u\sigma_u^4}\right)^{1/2}$ under our assumptions (but with time-invariant noise). Intuitively, because the noise-induced bias of the realized variance estimator ($M\sigma_u^2$) increases drastically with the number of observations, the number of observations which minimizes the unconditional MSE of realized variance is lower than the number of observations which minimizes its unconditional variance.

Importantly, if $\mathbf{Var}(\sigma_{t\varepsilon}^2) > (\mathbf{E}(\varepsilon^2))^2 = (\mathbf{E}(\sigma_{t\varepsilon}^2))^2$ under time-varying noise, then $M_2^* > M_1^*$. This last condition will be easily satisfied for our data. Specifically, we will find that $M_3^* > M_2^* > M_1^*$ and, of course, $R_{M_1^*}^2 > R_{M_2^*}^2 > R_{M_3^*}^2$. This result deserves attention. While a time-varying noise second moment can lead to relatively infrequent optimal sampling (M_1^*), optimizing the realized variance estimator's unconditional MSE (under the assumption of a constant noise variance), as implied by M_2^*, can be a superior strategy to focusing on the unconditional variance of realized variance (again under the assumption of a constant noise variance), as implied by M_3^*. This finding is particularly interesting since the latter choice would in fact be the optimal choice, from a forecasting standpoint, should the second moment of the noise (or the realized variance estimator's bias) be assumed to be time-invariant.

[6] Interestingly, this is the same rule obtained by Bandi and Russell (2003) in a different context, namely the finite sample MSE (variance) minimization of their proposed bias-corrected realized variance estimator.

5 Conditional vs. unconditional frequency choices

Rather than selecting one sampling frequency for all entries in the regressor vector (i.e., the vector of realized variance estimates), one could select a different optimal frequency for each entry/day. Bandi and Russell (2006a, 2008) use this approach in predicting variance on the basis of autoregressive (fractionally-integrated) models.

This section shows that whether the conditional approach has the potential to deliver superior forecasts than the unconditional approach described in the previous section depends on empirically verifiable conditions. Specifically, we consider the conditional finite sample MSE-based approximate rule in Bandi and Russell (2003, 2008), i.e.,

$$M_{2t}^{*} = \left(\frac{Q_t}{\sigma_{t\varepsilon}^4}\right)^{1/3}, \tag{8}$$

and compare it to the optimal, approximate unconditional rule in Eq. (5).

Theorem 2.

Define

$$\begin{aligned}
\mathbf{Var}_{M_1^*}(x_t) &= 2\left(\mathbf{Var}\left(\sigma_{t\varepsilon}^2\right)\right)^{1/3}\mathbf{E}(Q_t)^{2/3} + \left(\frac{\mathbf{E}(Q_t)}{\mathbf{Var}\left(\sigma_{t\varepsilon}^2\right)}\right)^{1/3}\mathbf{E}(2\theta_{t\varepsilon} - 3\sigma_{t\varepsilon}^4) \\
&\quad + \left(4\mathbf{E}(\sigma_{t\varepsilon}^2 V_t) - \mathbf{E}(\theta_{t\varepsilon}) + 2\mathbf{E}\left(\sigma_{t\varepsilon}^4\right)\right) \\
&\quad + \mathbf{Var}(V_t) + \left(\frac{\mathbf{E}(Q_t)}{\mathbf{Var}\left(\sigma_{t\varepsilon}^2\right)}\right)^{2/3}\left(\mathbf{Var}\left(\sigma_{t\varepsilon}^2\right)\right)
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{Var}_{M_{2t}^*}(x_t) &= 2\mathbf{E}\left(\left(\sigma_{t\varepsilon}^2\right)^{2/3}\right)\mathbf{E}\left(Q_t^{2/3}\right) + \mathbf{E}\left(\left(\frac{Q_t}{\sigma_{\varepsilon t}^4}\right)^{1/3}(2\theta_{t\varepsilon} - 3\sigma_{t\varepsilon}^4)\right) \\
&\quad + \left(4\mathbf{E}(\sigma_{t\varepsilon}^2 V_t) - \mathbf{E}(\theta_{t\varepsilon}) + 2\mathbf{E}\left(\sigma_{t\varepsilon}^4\right)\right) \\
&\quad + \mathbf{Var}(V_t) + \mathbf{Var}\left((Q_t)^{1/3}\left(\sigma_{t\varepsilon}^2\right)^{1/3}\right) + 2\mathbf{Cov}\left(V_t, (Q_t)^{1/3}\left(\sigma_{t\varepsilon}^2\right)^{1/3}\right).
\end{aligned}$$

If

$$\frac{\left(\mathbf{Var}(V_t) + \mathbf{Cov}\left(V_t, (Q_t)^{1/3}\left(\sigma_{t\varepsilon}^2\right)^{1/3}\right)\right)^2}{\mathbf{Var}_{M_{2t}^*}(x_t)} > \frac{(\mathbf{Var}(V_t))^2}{\mathbf{Var}_{M_1^*}(x_t)}, \tag{9}$$

then

$$R^2_{M_{2t}^*} > R^2_{M_1^*}.$$

Proof.

See Appendix.

Remark 5

The statement in Theorem 2 highlights the moment condition affecting the preferability of an approximate conditional rule versus an approximate unconditional rule (see Eq. (9)). Leaving aside issues related to estimation uncertainty, the inequality can be easily evaluated empirically by using the methods described in Remark 2 above. The condition is easily satisfied for our data.

Similarly, we can provide a statement for *exact* conditional and unconditional rules. Specifically, one could compare

$$\frac{(\mathbf{Var}(V_t))^2}{\mathbf{Var}_{M_1}(x_t)} \tag{10}$$

with

$$\begin{aligned}\mathbf{Var}_{M_1}(x_t) \quad = \quad & \frac{2}{M_1}\mathbf{E}(Q_t) + (2\mathbf{E}(\theta_{t\varepsilon}) - 3\mathbf{E}(\sigma^4_{t\varepsilon}))M_1 \\ & + \left(4\mathbf{E}(\sigma^2_{t\varepsilon}V_t) - \mathbf{E}(\theta_{t\varepsilon}) + 2\mathbf{E}\left(\sigma^4_{t\varepsilon}\right)\right) \\ & + \mathbf{Var}(V_t) + (M_1)^2\,\mathbf{Var}(\sigma^2_{t\varepsilon})\end{aligned}$$

to

$$\frac{\left(\mathbf{Var}(V_t) + \mathbf{Cov}\left(V_t, M_{2t}\sigma^2_{t\varepsilon}\right)\right)^2}{\mathbf{Var}_{M_{2t}}(x_t)} \tag{11}$$

with

$$\begin{aligned}\mathbf{Var}_{M_{2t}}(x_t) \quad = \quad & 2\mathbf{E}\left(\frac{Q_t}{M_{2t}}\right) + \mathbf{E}\left(M_{2t}(2\theta_{t\varepsilon} - 3\sigma^4_{t\varepsilon})\right) \\ & + \left(4\mathbf{E}(\sigma^2_{t\varepsilon}V_t) - \mathbf{E}(\theta_{t\varepsilon}) + 2\mathbf{E}\left(\sigma^4_{t\varepsilon}\right)\right) \\ & + \mathbf{Var}(V_t) + \mathbf{Var}\left(M_{2t}\sigma^2_{t\varepsilon}\right) + 2\mathbf{Cov}\left(V_t, M_{2t}\sigma^2_{t\varepsilon}\right),\end{aligned}$$

where M_1 is the exact R^2-optimal number of observations (from Theorem 1) and M_{2t} is the exact (conditional) MSE-optimal number of observations from Bandi and Russell (2003, 2008). If Eq. (11) is larger than Eq. (10), then $R^2_{M_{2t}} > R^2_{M_1}$. Naturally, this new inequality is slightly harder to verify than the inequality in the theorem. Its verification requires solution of $n+1$, where n is the number of days in the sample, optimization problems to compute the relevant M's (i.e., M_1 and M_{2t} with $t = 1, ..., n$).

6 Forecasting regressions in practise

This section examines the implications of theory with data. We use SPIDERS (Standard and Poor's depository receipts) mid-quotes on the NYSE.[7] We remove quotes whose associated price changes and/or spreads are larger than 10%.

In order to render the regressions feasible (i.e., to evaluate the regressand V_{t+1}), we employ flat-top kernels as advocated by Barndorff-Nielsen et al. (2008). Write

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

where $\widehat{\gamma}_s = \sum_{j=1}^{M} r_{t+j\delta} r_{t+(j-s)\delta}$ with $s = -q, ..., q$, $w_s = k\left(\frac{s-1}{q}\right)$ and $k(.)$ is a function on $[0,1]$ satisfying $k(0) = 1$ and $k(1) = 0$. The well-known Bartlett kernel ($k(x) = 1-x$), the cubic kernel ($k(x) = 1-3x^2+2x^3$), and the modified Tukey-Hanning kernel ($k(x) = \left(1 - \cos \pi(1-x)^2\right)/2$), among other functions, satisfy the conditions on $k(.)$. These estimators have favorable limiting properties under our price formation mechanism (Barndorff-Nielsen et al., 2008).[8] Furthermore, they have been shown to perform satisfactorily in practise (Bandi and Russell, 2010, and Bandi et al., 2008). Importantly, for each day in the sample, the estimators are unbiased under our assumptions. This is a useful property in that it guarantees (theoretically, at least) unbiasedness of the forecasts.[9] We optimize the performance of the estimators by using methods discussed in Bandi and Russell (2010). Specifically, for each day in the sample (i.e., "conditionally," using our previous terminology) we select the number of autocovariances λ in order to minimize the estimators' finite sample variance.[10]

We run regressions of $\widehat{V}_{t+1}^{BNHLS}$ on lagged realized variance. To fully capture the persistence properties of volatility, we run regressions of $\widehat{V}_{t+1}^{BNHLS}$ on five lags of realized variance. We consider two subsets of data, the full period 1/2002 - 3/2006 and the shorter 1/2004 - 3/2006 period. In all cases, we use 1,000 observations to estimate the model's parameters and forecast. We report 6 cases for the regressor(s):

1. Realized variance with M_{2t}^{*}.

[7] SPIDERS are shares in a trust which owns stocks in the same proportion as that found in the S&P 500 index. They 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. SPIDERS are exchange-traded funds. They can be redeemed for the underlying portfolio of assets. Equivalently, investors have the right to obtain newly issued SPIDERS shares from the fund company in exchange for a basket of securities reflecting the SPIDERS' portfolio.

[8] If $q \propto M^{2/3}$, the estimators are consistent and converge to an asymptotic mixed normal distribution at speed $M^{1/6}$. The additional requirements $k'(0) = 0$ and $k'(1) = 0$, combined with $q \propto M^{1/2}$, yield a faster rate of convergence ($M^{1/4}$) to the estimators' mixed normal distribution. The cubic kernel and the modified Tukey-Hanning kernel satisfy the extra requirements.

[9] In general, of course, how to optimally trade-off bias and variance of the forecasts depends on the adopted loss function. Since we are simply using kernel estimates to make the regressions feasible by empirically evaluating the regressand, it seems natural to employ unbiased estimators with favorable variance properties.

[10] Other estimators, such as the two-scale estimator of Zhang et al. (2005) and the multi-scale estimator of Zhang (2006), also have favorable properties and could be used.

2. Realized variance with a conditional version of M_3^*, i.e., $M_{3t}^* = \left(\frac{2Q_t}{2\theta_{t\varepsilon} - 3\sigma_{t\varepsilon}^4}\right)^{1/2}$.

3. Realized variance with an unconditional choice of M obtained by least-squares minimization (jointly with estimation of the regression's parameters), M_4^*.

4. Realized variance with $\overline{M}_{2t}^*$ (a unique M obtained by averaging the M_{2t}^*s across 1,000 observations - the same observations used for forecasting).

5. Realized variance with M_3^*.

6. Realized variance with $\overline{M}_{3t}^*$ (a unique M obtained by averaging the M_{3t}^*s across 1,000 observations - the same observations used for forecasting).

We begin by focusing on *unconditional* choices ((3) through (6)). The one-step estimator in (3) performs better than the estimator obtained by averaging conditional MSE-optimal frequencies in (4). In turn, the estimator in (4) outperforms the estimator (in (5)) obtained by choosing the unconditional optimal (in a variance sense) frequency. Notice that the one-step least-squares estimator is the estimator which empirically maximizes the regression's R^2 jointly with the regression's parameters. In other words, this estimator represents the least-squares solution to frequency optimization in the context of forecasting. Under our assumptions, this estimator should provide a similar answer as the exact theoretical solution M_1. Hence, its superiority is not surprising. It is also not surprising that taking the time-varying noise into account through an MSE criterion (as implied by (4)) appears to be a superior strategy than simply focusing on the unconditional variance of the regressor under the assumption of a time-invariant noise (as implied by (5)).

We find that $M_4^* < \overline{M}_{2t}^* < M_3^*$. In other words, $q_4^* > \overline{q}_{2t}^* > q_3^*$, where q indicates the average number of high-frequency observations to be skipped in the computation of the corresponding realized variance estimator. These values are reported in the tables. For the case of the flat-top Bartlett kernel in Table 1, for example, $q_4^* \approx 123$, $\overline{q}_{2t}^* \approx 74$, and $q_3^* \approx 55$. Since, empirically, $\mathbf{Var}(\sigma_{t\varepsilon}^2)(= 5.19e-14) > \left(\mathbf{E}(\sigma_{t\varepsilon}^2)\right)^2 (= 1.6e-14)$, the ranking of M values is fully consistent with theory. Differently put, it is consistent with the ranking reported in Section 3 above, namely $M_1^* < M_2^* < M_3^*$. This said, M_4^* and $\overline{M}_{2t}^*$ are not exactly equal to M_1^* and M_2^*, respectively. However, we expect them to be very close to M_1^* and M_2^*. While, as noted, M_4^* is the empirical R^2-maximizing number of observations, M_1^* is a useful approximation to it highlighting the main determinants of the optimal frequency in the time-varying noise case. In addition, $\overline{M}_{2t}^*$ and M_2^* are not identical objects due to Jensen's inequality-type arguments.

In our sample, averaging the conditional, optimal (in a variance sense) frequencies (as in (6)) is an inferior strategy to using the unconditional, variance-optimal frequency (as in (5)) advocated by ABM (2010) and GS (2010).

Turning to *conditional* choices, we find that using conditional MSE-optimal frequencies as in (1) and conditional variance-optimal frequencies as in (2) performs better than selecting only one frequency (even if it is the R^2-optimal frequency in (3)) for all realized variance entries. Importantly, as stressed earlier, the moment condition in Theorem 2 is easily satisfied in our sample. Hence, these results nicely conform with the implications of theory too.

Our findings are consistent across regressands (i.e., choices of kernels) and sample periods. Statistically, we find that (*i*) the forecast MSEs are always different from zero, (*ii*) a chi-squared test of the null hypothesis of equal MSEs across forecasting models is easily rejected, and (*iii*) pairwise t-tests of the null of equal MSEs between conditional and unconditional frequency choices are generally rejected in the low variance sample period 1/2004 - 3/2006 (Tables 4, 5, and 6).

7 Alternative robust regressors

We have shown that the presence of a time-varying bias should lead to careful consideration when forecasting variance using realized variance. Not only is this type of bias not absorbed by the regression's intercept in general, but it can lead to unconditional choices of sampling frequency that are lower, rather than higher as in the constant bias case, than the realized variance estimator's unconditional MSE-optimal choice (under the assumption of a constant bias). We have also found that conditional choices can be beneficial in practise.

One could of course forecast variance using alternative, theoretically more robust (to microstructure noise) regressors than the classical realized variance estimator. In this section we employ the same regressors that were used earlier as regressands, namely flat-top Bartlett kernels, flat-top cubic kernels, and flat-top modified Tukey-Hanning kernels. When using the flat-top Bartlett kernel as a regressand, for instance, we also use it as a regressor. The forecasting set-up is the same as that in the previous section. We consider 5 lags for the regressor to capture volatility persistence effectively. We employ 1,000 observations to estimate the model's parameters and forecast. We now report 3 cases for the regressor(s):

1. Kernel estimates with a number of autocovariances chosen conditionally (for each day) in order to minimize the finite sample MSE (variance, in this case) of the estimator as suggested by Bandi and Russell (2010).

2. Kernel estimates with a number of autocovariances chosen unconditionally in order to minimize least-squares (i.e., jointly with the model's parameters).

3. Kernel estimates with a number of autocovariances chosen unconditionally as the average of the conditional MSE(variance)-optimal choices in Bandi and Russell (2010).

As earlier, we begin with *unconditional* choices. Even in the case of estimators with favorable (theoretically) bias properties in the presence of noise, there appears to be scope for taking the documented time-variation in the noise second moment into account from a forecasting perspective. Unconditional autocovariance/bandwidth choices which empirically maximize the R^2 of the regression perform better than unconditional variance-optimal bandwidth choices (as in (3)). As thoroughly discussed by ABM (2010) and GS (2010), the latter would be optimal should the noise-induced bias be constant (or absent, as is the case for the flat-top kernels in theory). It is interesting to notice that the average number of autocovariances that is selected by the one-step R^2-optimal procedure is larger (around 13) then the average number of autocovariances selected by the variance-optimal procedure (around 5). The divergence between the unconditional choices in (2) and (3) is potentially due to a time-varying (noise-induced) bias component that is left in the kernel estimates despite their favorable (theoretically) finite sample bias properties. Qualitatively and quantitatively, this result is reminiscent of findings in Bandi and Russell (2010). There it was shown that the (conditional) finite sample MSE-based optimization of the promising two-scale estimator results in a larger number of autocovariances than the finite sample MSE-based optimization of the class of flat-top kernels. This outcome was due to the presence of a finite sample bias component in the case of the former and the need to reduce it for the purpose of MSE minimization.[11]

Turning to *conditional* bandwidth choices, choosing the autocovariances to optimize the finite sample variance of the estimator (as in (1)) for each entry in the regressor vector outperforms, in our sample, the "best" unconditional choice in (2). This is, again, consistent with our findings in the realized variance case.

The use of robust regressors leads, as expected, to gains over the use of the realized variance estimator. Interestingly, however, these gains can be fairly small in practise. Consider the case $\widehat{V}_{t+1}^{BNHLS}$ = Bartlett and $\widehat{V}_t$ = realized variance vs. the case $\widehat{V}_{t+1}^{BNHLS}$ = Bartlett and $\widehat{V}_t^{BNHLS}$ = Bartlett for the optimal R^2-based unconditional choice. The corresponding forecast MSE values are $1.08e-09$ and $9.92e-10$. When using cubic kernels, the values are $1.1819e-09$ and $1.1814e-09$, respectively. They are $1.0986e-09$ and $1.0753e-09$ in the modified Tukey-Hanning kernel case.

8 Extending the framework: dependence between noise variance and true price variance

Sound economic reasoning suggests the potential for non-negligible cross-sectional dependence between noise variance and true price variance. For instance, higher uncertainty about the asset's true value (as represented by a higher V_t) implies higher likelihood of adverse price moves and, hence, higher inventory risk for the market maker. Similarly, higher uncertainty about the true value of the asset increases the market maker's

[11] The proposed analogy should only be regarded as suggestive in that the finite sample bias of the two-scale estimator simply depends on the moments of the underlying price process and not on the moments of the noise contaminations.

risk of transacting with traders with superior information. Both risks ought to be compensated. Generally, other things equal, higher price variance translates into larger bid-ask spreads set by the market maker and hence larger noise variances. Bandi and Russell (2006a) provide discussions about the economic reasons underlying the *cross-sectional dependence* between true price variance and noise variance as well as empirical evidence for the S&P100 stocks.

A related issue has to do with the *time-series dependence* between true price variance and noise variance. From a forecasting perspective, if $\mathbf{Cov}(V_t, \sigma^2_{tu}) \neq 0$, the R^2-optimal rule to sample realized variance should be modified.

Theorem 3.

Consider the regression in Eq. (1). *If* $\mathbf{Cov}(V_t, \sigma^2_{tu}) \neq 0$, *then*

$$M_1^{\blacktriangle} = \arg\max R^2_M = \arg\max \frac{(\mathbf{Var}(V_t) + M\mathbf{Cov}(V_t, \sigma^2_{t\varepsilon}))^2}{\mathbf{Var}_M(x_t)}, \tag{12}$$

where

$$\begin{aligned}\mathbf{Var}_M(x_t) \quad = \quad & \frac{2}{M}\mathbf{E}(Q_t) + M\mathbf{E}(2\theta_{t\varepsilon} - 3\sigma^4_{t\varepsilon}) + \left(4\mathbf{E}(\sigma^2_{t\varepsilon}V_t) - \mathbf{E}(\theta_{t\varepsilon}) + 2\mathbf{E}\left(\sigma^4_{t\varepsilon}\right)\right) \\ & + \mathbf{Var}\left(V_t\right) + M^2\mathbf{Var}\left(\sigma^2_{t\varepsilon}\right) + 2M\mathbf{Cov}(V_t, \sigma^2_{t\varepsilon}).\end{aligned}$$

Proof.

See Appendix.

Given consistent daily estimates of $\sigma^2_{t\varepsilon}$ and preliminary estimates of V_t (obtained as described in Remark 2 above), the additional input $\mathbf{Cov}(V_t, \sigma^2_{t\varepsilon})$ can be readily evaluated. Numerical maximization of R^2_M can then be easily conducted.

Naturally, the least-squares solution M_4^* (above) accounts for the presence of a non-zero covariance between V_t and $\sigma^2_{t\varepsilon}$ directly. Under our assumptions, this solution should be close to $M_1^{\blacktriangle}$. Having made this point, a thorough study of the dynamic properties of the bi-variate system $(V_t, \sigma^2_{t\varepsilon})$, and their implications for forecasting integrated variance, is an important topic of research better left for future work.

9 A general solution to the unconditional problem: *joint* least-squares optimization

Our previous discussion used a conventional price formation mechanism (extended to allow for time-varying noise second moments) as well as a simple, but commonly-employed, forecasting model. Both choices were intended to obtain closed-form implications for important determinants of the optimal bandwidth/frequency choice when noise is time-varying. In this setting, we have shown that there are theoretical reasons for choosing unconditional sampling frequencies (bandwidths) which are lower (larger) than those that would be optimally chosen in linear forecasting models when time-variation in the second moment of the noise is unaccounted for. Our empirical work confirmed this implication of theory.

The solutions we derived were, of course, intended to minimize the mean-squared forecast error of the assumed forecasting regressions. Importantly, this optimization may be conducted without imposing potentially strong assumptions on the noise properties and in the context of any variance forecasting model relying on realized variance measures. Coherently with our general logic as laid out in Section 3 (as well as our M_4^* choice in Section 6), here we propose to evaluate the optimal frequency/bandwidth *jointly* with the model's parameters.

Let $V_{t+1} = f(V_t, V_{t-1}, ..., V_{t-(K-1)}|\theta) + \xi_{t+1}$ with $\mathbf{E}(\xi_{t+1}|\mathfrak{F}_t) = 0$ and define $V_{t+1} = f(\widehat{V}_{\widehat{\phi},t}, \widehat{V}_{\widehat{\phi},t-1}, ..., \widehat{V}_{\widehat{\phi},t-(K-1)}|\widehat{\theta}) + \widehat{\xi}_{t+1}$, for some function $f(.)$ and a specific number of lags $K > 0$. The least-squares solution to the joint forecasting/sampling problem is given by:

$$\arg\min_{\phi,\theta} \sum_{t=1}^{T} \widehat{\xi}_{t+1} = \arg\min_{\phi,\theta} \sum_{t=1}^{T} [V_{t+1} - f(\widehat{V}_{\phi,t}, \widehat{V}_{\phi,t-1}, ..., \widehat{V}_{\phi,t-(K-1)}|\theta)]^2,$$

where $\widehat{V}_{\phi,t+h}$ is a realized variance measure and ϕ is either a frequency (in the case of realized variance) or a bandwidth (in the case of kernel estimates).

Under our assumption on the noise, if $f(.)$ is an AR(1) model, the solution to the (non-linear) least-squares problem coincides with the theoretical solution in Section 3. More generally, this solution is robust to potential noise dependence, dependence between the noise and the equilibrium price process, as well as dependence between the equilibrium price variance and the noise variance, as discussed in the previous section. Moreover, the forecasting model may be richer than an autoregression of any order. For instance, the joint approach readily applies to the MIDAS regressions in GS (2010), *inter alia*, to the heterogeneous autoregressive regressions (HAR) of Corsi (2009), possibly with leverage effects and jumps as in Corsi and Renò (2010), and to the HEAVY specifications in Shephard and Sheppard (2009), among other approaches.

We note that the specification $f(\widehat{V}_{\phi,t}, \widehat{V}_{\phi,t-1}, ..., \widehat{V}_{\phi,t-(K-1)}|\theta)$ can be viewed as a mixed parameter model in that ϕ is naturally defined over a discrete set. Under conventional assumptions (see, e.g., Choirat and Seri, 2001, and Ryu, 1999), it is readily shown that $\mathbf{E}(\widehat{\theta}_i - \theta_{i0})^2 \leq \zeta T^{-1}$, for $0 \leq i \leq K-1$ and some $\zeta > 0$, and $\mathbf{E}(\widehat{\phi} - \phi_0)^2 \leq \rho^T$ for some $0 < \rho < 1$. In other words, the discrete parameter vector $\widehat{\phi}$ converges (in mean square) to its theoretical counterpart at a faster (exponential) rate than the classical root T rate. In the context of a simple AR(1) model, the values $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_0)$ are a slope and an intercept estimate, respectively, consistent for the pseudo-true parameters β_0 in Eq. (3) and $\alpha_0 = \mathbf{E}(V_{t+1}) - \beta_0 \mathbf{E}(V_t)$ (c.f., Section 3). As shown, β_0 and α_0 do not coincide with the parameters of the true autoregression but are such that $\beta_0 < \beta$ and $\alpha_0 > \alpha$. In addition, $\widehat{\phi}$ is consistent for the value ϕ_0 which minimizes the variance of the estimated regressor (as discussed in Section 3 and shown in Theorem 1). The same logic applies to the general specification $f(\widehat{V}_{\phi,t}, \widehat{V}_{\phi,t-1}, ..., \widehat{V}_{\phi,t-(K-1)}|\theta)$ for which $\widehat{\theta}$ is a consistent estimate of the attenuated pseudo-true parameter vector $\theta_0 \neq \theta$ and $\widehat{\phi}$ is consistent (at the accelerated rate $\rho^{-\frac{1}{2}T}$) for the value ϕ_0 which minimizes the regressor's variance.

10 Conclusions

The second moment of market microstructure noise is time-varying. We study the impact of this time variation on variance forecasting. In the context of linear volatility forecasting models, we find the need for lower sampling frequencies (in the case of realized variance) and larger bandwidth choices (in the case of kernel estimators of integrated variance) than required when microstructure noise is assumed to be present but its variability is unaccounted for. We also show that frequency/bandwidth choices which adapt to the moments of the true price and noise components (named "conditional" in the text) have the potential to outperform (from an R^2 or forecast MSE standpoint) their optimal unconditional counterparts.

Importantly, the goal of this paper is *not* to advocate a specific variance forecasting model. Choosing an optimal lag structure in the relevant forecasting regressions, as well as enlarging the information set by allowing for additional predictors (aside from lagged variance), are, among other extensions, important issues beyond the scopes of the present paper. Our goal is to use a well-understood price structure, as well as a classical loss function amenable to the derivation of clear theoretical implications, to highlight aspects of the volatility forecasting problem in the presence of market microstructure noise which we regard as important. Specifically, we show that (*1*) accounting for the time-varying nature of the noise moments, (*2*) deriving unconditional solutions by jointly selecting the frequency/bandwidth and the model's parameters, and (*3*) allowing for conditional frequency/bandwidth choices may be very beneficial in practise. The evaluation of richer forecasting models and alternative (statistical and economic) loss functions is better left for future work.

11 Appendix

Proof of Theorem 1. Consider the d.g.p. $V_{t+1} = \alpha + \beta V_t + \xi_{t+1}$, where ξ_{t+1} is a forecast error uncorrelated with time t information. We run a regression of V_{t+1} on $x_t = V_t + U_t + (x_t - V_t - U_t) = V_t + U_t + \eta_t$, where $U_t = \mathbf{E}\left(\sum_{j=1}^{M} \varepsilon_j^2 | \sigma, \sigma_u\right) = M\sigma_{t\varepsilon}^2$, and $\eta_t = \left(\sum_{j=1}^{M} r_j^2 - V_t\right) + \sum_{j=1}^{M} r_j \varepsilon_j + \left(\sum_{j=1}^{M} \varepsilon_j^2 - U_t\right)$. Notice that $\mathbf{E}(\eta_t|\sigma, \sigma_u) = 0$. Now, consider $R^2 = \frac{\mathbf{Cov}^2(V_{t+1}, x_t)}{\mathbf{Var}(V_{t+1})\mathbf{Var}(x_t)}$. The (square root of the) numerator can be expressed as $\mathbf{Cov}(V_{t+1}, x_t) = \mathbf{Cov}(\alpha + \beta V_t + \xi_{t+1}, V_t + U_t + \eta_t) = \beta \mathbf{Var}(V_t) + \beta \mathbf{Cov}(V_t, \eta_t)$ since $\mathbf{Cov}(V_t, U_t) = 0$. But,

$$
\begin{aligned}
\mathbf{Cov}(V_t, \eta_t) &= \mathbf{E}(V_t \eta_t) - \mathbf{E}(V_t)\mathbf{E}(\eta_t) \\
&= \mathbf{E}(V_t \eta_t) \\
&= \mathbf{E}\left(\mathbf{E}(V_t \eta_t | \sigma, \sigma_u)\right) \\
&= \mathbf{E}(V_t \mathbf{E}(\eta_t | \sigma, \sigma_u)) \\
&= 0.
\end{aligned}
$$

Hence, $\min_M \mathbf{Var}(x_t) \Rightarrow \max_M R^2$. By the law of total variance and Theorem 4 in Bandi and Russell (2008), write

$$
\begin{aligned}
\mathbf{Var}(x_t) &= \mathbf{E}(\mathbf{Var}(x_t|\sigma, \sigma_u)) + \mathbf{Var}(\mathbf{E}(x_t|\sigma, \sigma_u)) \\
&= \mathbf{E}\left(\frac{2}{M} Q_t\right) + \mathbf{E}(M(2\theta_{t\varepsilon} - 3\sigma^4_{t\varepsilon})) + \left(4\mathbf{E}(\sigma^2_{t\varepsilon} V_t) - \mathbf{E}(\theta_{t\varepsilon}) + 2\mathbf{E}\left(\sigma^4_{t\varepsilon}\right)\right) \\
&\quad + \mathbf{Var}\left(V_t + M\sigma^2_{t\varepsilon}\right) \\
&= \frac{2}{M}\mathbf{E}(Q_t) + M\mathbf{E}(2\theta_{t\varepsilon} - 3\sigma^4_{t\varepsilon}) + \left(4\mathbf{E}(\sigma^2_{t\varepsilon} V_t) - \mathbf{E}(\theta_{t\varepsilon}) + 2\mathbf{E}\left(\sigma^4_{t\varepsilon}\right)\right) \\
&\quad + \mathbf{Var}(V_t) + M^2 \mathbf{Var}\left(\sigma^2_{t\varepsilon}\right), \qquad (13)
\end{aligned}
$$

since, again, $\mathbf{Cov}(V_t, U_t) = 0$.■

Proof of Theorem 2. Consider Eq. (13). Plugging in the approximate unconditional rule $M_1^* = \left(\frac{\mathbf{E}(Q_t)}{\mathbf{Var}\left(\sigma^2_{t\varepsilon}\right)}\right)^{1/3}$ we obtain

$$
\begin{aligned}
\mathbf{Var}_{M_1^*}(x_t) &= \frac{2}{\left(\frac{\mathbf{E}(Q_t)}{\mathbf{Var}\left(\sigma^2_{t\varepsilon}\right)}\right)^{1/3}} \mathbf{E}(Q_t) + \mathbf{E}\left(\left(\frac{\mathbf{E}(Q_t)}{\mathbf{Var}\left(\sigma^2_{t\varepsilon}\right)}\right)^{1/3} (2\theta_{t\varepsilon} - 3\sigma^4_{t\varepsilon})\right) \\
&\quad + \left(4\mathbf{E}(\sigma^2_{t\varepsilon} V_t) - \mathbf{E}(\theta_{t\varepsilon}) + 2\mathbf{E}\left(\sigma^4_{t\varepsilon}\right)\right) \\
&\quad + \mathbf{Var}(V_t) + \left(\frac{\mathbf{E}(Q_t)}{\mathbf{Var}\left(\sigma^2_{t\varepsilon}\right)}\right)^{2/3} \left(\mathbf{Var}\left(\sigma^2_{t\varepsilon}\right)\right) \\
&= 2\mathbf{Var}\left(\sigma^2_{t\varepsilon}\right)^{1/3} \mathbf{E}(Q_t)^{2/3} + \left(\frac{\mathbf{E}(Q_t)}{\mathbf{Var}\left(\sigma^2_{t\varepsilon}\right)}\right)^{1/3} \mathbf{E}(2\theta_{t\varepsilon} - 3\sigma^4_{t\varepsilon}) \\
&\quad + \left(4\mathbf{E}(\sigma^2_{t\varepsilon} V_t) - \mathbf{E}(\theta_{t\varepsilon}) + 2\mathbf{E}\left(\sigma^4_{t\varepsilon}\right)\right) \\
&\quad + \mathbf{Var}(V_t) + \mathbf{E}(Q_t)^{2/3} \mathbf{Var}\left(\sigma^2_{t\varepsilon}\right)^{1/3}. \qquad (14)
\end{aligned}
$$

Hence, $R^2_{M_1^*} = \frac{\beta^2 (\mathbf{Var}(V_t))^2}{\mathbf{Var}(V_{t+1}) \mathbf{Var}_{M_1^*}(x_t)}$. Using the conditional rule in Bandi and Russell (2003, 2008):

$$
\begin{aligned}
\mathbf{Var}_{M_{2t}^*}(x_t) &= \mathbf{E}\left(\frac{2}{\left(\frac{Q_t}{\sigma_{t\varepsilon}^4}\right)^{1/3}}Q_t\right) + \mathbf{E}\left(\left(\frac{Q_t}{\sigma_{t\varepsilon}^4}\right)^{1/3}(2\theta_{t\varepsilon} - 3\sigma_{t\varepsilon}^4)\right) \\
&+ \left(4\mathbf{E}(\sigma_{t\varepsilon}^2 V_t) - \mathbf{E}(\theta_{t\varepsilon}) + 2\mathbf{E}\left(\sigma_{t\varepsilon}^4\right)\right) + \mathbf{Var}\left(V_t + M\sigma_{t\varepsilon}^2\right) \\
&= 2\mathbf{E}\left(\left(\sigma_{t\varepsilon}^2\right)^{2/3}(Q_t)^{2/3}\right) + \mathbf{E}\left(\left(\frac{Q_t}{\sigma_{t\varepsilon}^4}\right)^{1/3}(2\theta_{t\varepsilon} - 3\sigma_{t\varepsilon}^4)\right) \\
&+ \left(4\mathbf{E}(\sigma_{t\varepsilon}^2 V_t) - \mathbf{E}(\theta_{t\varepsilon}) + 2\mathbf{E}\left(\sigma_{t\varepsilon}^4\right)\right) + \mathbf{Var}\left(V_t + \left(\frac{Q_t}{\sigma_{t\varepsilon}^4}\right)^{1/3}\sigma_{t\varepsilon}^2\right) \\
&= 2\mathbf{E}\left(\left(\sigma_{t\varepsilon}^2\right)^{2/3}\right)\mathbf{E}(Q_t^{2/3}) + \mathbf{E}\left(\left(\frac{Q_t}{\sigma_{t\varepsilon}^4}\right)^{1/3}(2\theta_{t\varepsilon} - 3\sigma_{t\varepsilon}^4)\right) \\
&+ \left(4\mathbf{E}(\sigma_{t\varepsilon}^2 V_t) - \mathbf{E}(\theta_{t\varepsilon}) + 2\mathbf{E}\left(\sigma_{t\varepsilon}^4\right)\right) \\
&+ \mathbf{Var}\left(V_t\right) + \mathbf{Var}\left((Q_t)^{1/3}\left(\sigma_{t\varepsilon}^2\right)^{1/3}\right) + 2\mathbf{Cov}(V_t, (Q_t)^{1/3}\left(\sigma_{t\varepsilon}^2\right)^{1/3}).
\end{aligned}
$$

Thus, $R^2_{M_{2t}^*} = \frac{\beta^2(\mathbf{Var}(V_t)+\mathbf{Cov}(V_t,U_t))^2}{\mathbf{Var}(V_{t+1})\mathbf{Var}_{M_{2t}^*}(x_t)} = \frac{\beta^2\left(\mathbf{Var}(V_t)+\mathbf{Cov}\left(V_t,\left(\frac{Q_t}{\sigma_{t\varepsilon}^4}\right)^{1/3}\sigma_{t\varepsilon}^2\right)\right)^2}{\mathbf{Var}(V_{t+1})\mathbf{Var}_{M_{2t}^*}(x_t)} = \frac{\beta^2\left(\mathbf{Var}(V_t)+\mathbf{Cov}\left(V_t,(Q_t)^{1/3}\left(\sigma_{t\varepsilon}^2\right)^{1/3}\right)\right)^2}{\mathbf{Var}(V_{t+1})\mathbf{Var}_{M_{2t}^*}(x_t)}$.

■

Proof of Theorem 3. If $\mathbf{Cov}(V_t, \sigma_{tu}^2) \neq 0$, then $\mathbf{Cov}\left(V_{t+1}, x_t\right) = \mathbf{Cov}(\alpha + \beta V_t + \xi_{t+1}, V_t + U_t + \eta_t) = \beta\mathbf{Var}(V_t) + \beta M\mathbf{Cov}(V_t, \sigma_{t\varepsilon}^2)$ from the proof of Theorem 1. In addition,

$$
\begin{aligned}
\mathbf{Var}(x_t) &= \frac{2}{M}\mathbf{E}(Q_t) + \mathbf{E}(M(2\theta_{t\varepsilon} - 3\sigma_{t\varepsilon}^4)) + \left(4\mathbf{E}(\sigma_{t\varepsilon}^2 V_t) - \mathbf{E}(\theta_{t\varepsilon}) + 2\mathbf{E}\left(\sigma_{t\varepsilon}^4\right)\right) \\
&+ \mathbf{Var}\left(V_t\right) + M^2\mathbf{Var}\left(\sigma_{t\varepsilon}^2\right) + 2M\mathbf{Cov}(V_t, \sigma_{t\varepsilon}^2). \qquad (15)
\end{aligned}
$$

Hence,

$$
R^2 = \frac{\beta^2}{\mathbf{Var}(V_{t+1})}\frac{(\mathbf{Var}(V_t) + M\mathbf{Cov}(V_t, \sigma_{t\varepsilon}^2))^2}{\mathbf{Var}(x_t)}
$$

with $\mathbf{Var}(x_t)$ defined in Eq. (15).

■

References

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

[2] Aït-Sahalia, Y., P. Mykland, and L. Zhang (2010). Ultra high-frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics, forthcoming.*

[3] Andersen, T.G., T. Bollerslev, and N. Meddahi (2004). Analytic evaluation of volatility forecasts. *International Economic Review 45*, 1079-1110.

[4] Andersen, T.G., T. Bollerslev, and N. Meddahi (2010). Realized volatility forecasting and market microstructure noise. *Journal of Econometrics*, *forthcoming.*

[5] Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica 71*, 579-625.

[6] Bandi, F.M., and J.R. Russell (2003). Microstructure noise, realized volatility, and optimal sampling. *Working paper.*

[7] Bandi, F.M., and J.R. Russell (2006a). Separating microstructure noise from volatility. *Journal of Financial Economics 79,* 655-692.

[8] Bandi, F.M., and J.R. Russell (2006b). Comment on Hansen and Lunde. *Journal of Business and Economic Statistics 24,* 167-173.

[9] Bandi, F.M., and J.R. Russell (2007). Volatility. In J. Birge and V. Linetski (Eds.) *Handbook of Financial Engineering.* Elsevier, North Holland.

[10] Bandi, F.M., and J.R. Russell (2008). Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies 75,* 339-369

[11] Bandi, F.M., and J.R. Russell (2010). Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations. *Journal of Econometrics, forthcoming.*

[12] Bandi, F.M., J.R. Russell, and C. Yang (2008). Realized volatility forecasting and option pricing. *Journal of Econometrics 147*, 34-46.

[13] Barndorff-Nielsen, O.E., and N. Shephard (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B 64*, 253-280.

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

[15] Barndorff-Nielsen, O.E., P. Hansen, A. Lunde and N. Shephard (2008). Designing realized kernels to measure ex-post variation of equity prices in the presence of noise. *Econometrica 76,* 1481-1536.

[16] Choirat, C. and R. Seri (2001). Estimation in discrete parameter models. *Working paper.*

[17] Corradi, V., W. Distaso, and N. Swanson (2006). Predictive inference for integrated volatility. *Working paper.*

[18] Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics 7*, 174-196.

[19] Corsi, F., and R. Renò (2010). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Working paper.*

[20] Ghysels, E. and A. Sinko (2006). Comment on Hansen and Lunde. *Journal of Business and Economic Statistics 24,* 192-194.

[21] Ghysels, E. and A. Sinko (2010). Volatility forecasting and microstructure noise. *Journal of Econometrics, forthcoming.*

[22] Hansen, P.R., and A. Lunde (2006). Realized variance and market microstructure noise (with discussions). *Journal of Business and Economic Statistics 24,* 127-161.

[23] Kalnina, I. and O. Linton (2008). Estimating quadratic variation consistently in the presence of correlated measurement error. *Journal of Econometrics 147,* 47-59.

[24] McAleer, M., and M. Medeiros (2008). Realized volatility: A review. *Econometric Reviews 27,* 10-45.

[25] Oomen, R.C.A. (2005). Properties of bias-corrected realized variance under alternative sampling schemes. *Journal of Financial Econometrics 3,* 555–577.

[26] Oomen, R.C.A. (2006). Properties of realized variance under alternative sampling schemes. *Journal of Business and Economic Statistics 24,* 219-237.

[27] Ryu, K (1999). Econometric Analysis of mixed parameter models. *Working paper.*

[28] Shephard, N. and K. Sheppard (2009). Realising the future: forecasting with high-frequency based (HEAVY) volatility models. *Journal of Applied Econometrics* 23, 197-231.

[29] Zhang, L., P. Mykland, and Y. Aït-Sahalia (2005). A tale of two time scales: determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association 100,* 1394-1411.

[30] Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. *Bernoulli 12,* 1019-1043.

[31] Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business and Economic Statistics 14*, 45-52.

Figures and Tables

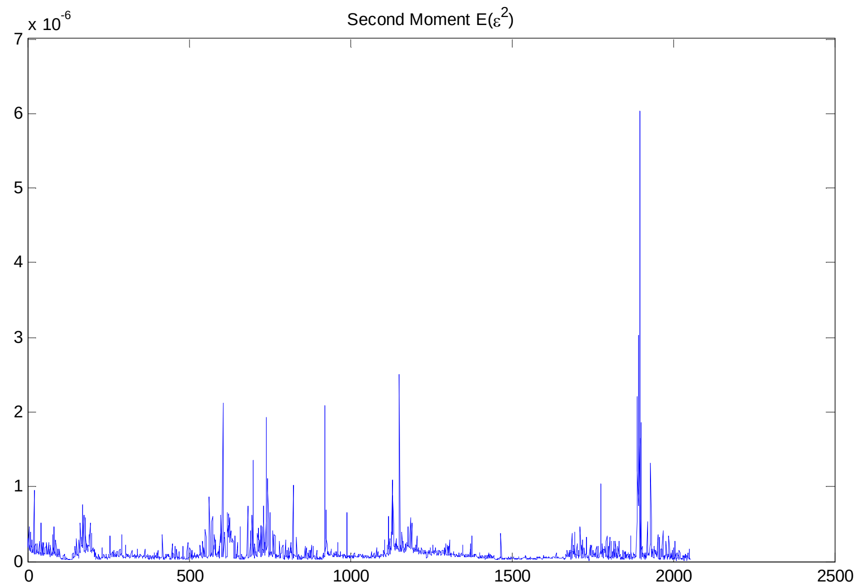

Figure 1. Microstructure noise second moment's estimates. We use Spiders mid-quotes on the NYSE. The sample period is 2002/1 - 2006/3.

LHS = FlatBart RHS = RV 5 lags
2002/1-2006/3

	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	162.326	**9.66E-10**	4.67E-05	5.88E-05	5.00E-05	4.93E-05
(1)	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.32E-05	0.281312	0.0989931	-0.0033682	0.0937042	0.045113
	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	117.478	**9.98E-10**	4.67E-05	5.80E-05	5.00E-05	4.29E-05
(2)	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.43E-05	0.318263	0.0556906	0.0192768	0.102851	0.00968516
	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	123.879	**1.08E-09**	4.67E-05	6.26E-05	5.00E-05	4.86E-05
(3)	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.77E-05	0.219259	0.0998483	0.0352042	0.0524899	0.0425611
	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
(4)	74.8889	**1.09E-09**	4.67E-05	6.18E-05	5.00E-05	4.51E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.94E-05	0.210582	0.0894445	0.0269597	0.0575391	0.0497621
	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	55.8889	**1.13E-09**	4.67E-05	6.25E-05	5.00E-05	4.57E-05
(5)	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.86E-05	0.212029	0.0925057	0.0296564	0.055968	0.0469369
	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	53.9468	**1.19E-09**	4.67E-05	6.26E-05	5.00E-05	4.54E-05
(6)	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.79E-05	0.21654	0.0936616	0.0285354	0.0555054	0.0496944

	(1)	(2)	(3)	(4)	(5)	(6)
MSE	9.66E-10	9.98E-10	1.08E-09	1.09E-09	1.13E-09	1.19E-09
HAC_std	2.34E-10	2.41E-10	2.20E-10	2.25E-10	2.10E-10	2.40E-10
t_MSE	4.1291	4.13497	4.91409	4.85042	5.37764	4.97761
	t12	t13	t14	t15	t25	
T	-0.24062	-1.32543	-1.54206	-1.66828	-1.02195	

Joint Chi-squared test: 50.41

Table 1. We report forecasting regressions of integrated variance (estimated using optimally-defined flat-top Bartlett kernels as in Bandi and Russell, 2010) on five lags of realized variance. The regressor (realized variance) is sampled using 6 methods described in the main text. We use Spiders mid-quotes on the NYSE. The sample period is 2002/1 - 2006/3. In all cases, 1,000 observations are employed to estimate the model parameters (a0 through a5) and forecast. The table reports means and standard deviations for both the regressand (true) and the volatility forecasts. It also reports the average number of observations to be skipped (q) according to each sampling rule. T-statistics for the individual MSEs, t-statistics for pair-wise tests of equal MSEs, and a joint Chi-squared test of equal MSEs across sampling methods are reported in the second panel.

LHS = FlatCubic RHS = RV 5 lags
2002/1-2006/3

(1)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	162.326	**1.05E-09**	4.85E-05	5.98E-05	5.19E-05	4.87E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.42E-05	0.287751	0.10447	-0.0050108	0.0934314	0.0368125
(2)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	117.478	**1.07E-09**	4.85E-05	5.90E-05	5.19E-05	4.23E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.50E-05	0.324704	0.0627989	0.0173676	0.102172	0.00247085
(3)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	117.969	**1.1819E-09**	4.85E-05	6.41E-05	5.19E-05	4.82E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.86E-05	0.222019	0.103153	0.0380231	0.0497926	0.0371127
(4)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	74.8889	**1.1871E-09**	4.85E-05	6.29E-05	5.19E-05	4.46E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	3.03E-05	0.21586	0.0941549	0.0257633	0.0545427	0.0457488
(5)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	55.8889	**1.22E-09**	4.85E-05	6.36E-05	5.19E-05	4.52E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.95E-05	0.217557	0.0973909	0.0281942	0.0531524	0.0425865
(6)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	53.9468	**1.30E-09**	4.85E-05	6.37E-05	5.19E-05	4.50E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.88E-05	0.222269	0.0983164	0.0269882	0.0528116	0.0453233

	(1)	(2)	(3)	(4)	(5)	(6)
MSE	1.05E-09	1.07E-09	1.1819E-09	1.1871E-09	1.22E-09	1.30E-09
HAC_std	2.78E-10	2.84E-10	2.60E-10	2.66E-10	2.50E-10	2.85E-10
t_MSE	3.77433	3.77673	4.57811	4.46129	4.88257	4.5744
	t12	t13	t14	t15	t25	
T	-0.1656	-1.4041	-1.50741	-1.57473	-1.01939	

Joint Chi-squared test: 43.78

Table 2. We report forecasting regressions of integrated variance (estimated using optimally-defined flat-top cubic kernels as in Bandi and Russell, 2010) on five lags of realized variance. The regressor (realized variance) is sampled using 6 methods described in the main text. We use Spiders mid-quotes on the NYSE. The sample period is 2002/1 - 2006/3. In all cases, 1,000 observations are employed to estimate the model parameters (a0 through a5) and forecast. The table reports means and standard deviations for both the regressand (true) and the volatility forecasts. It also reports the average number of observations to be skipped (q) according to each sampling rule. T-statistics for the individual MSEs, t-statistics for pair-wise tests of equal MSEs, and a joint Chi-squared test of equal MSEs across sampling methods are reported in the second panel.

LHS = FlatTukey RHS = RV 5 lags
2002/1-2006/3

(1)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	162.326	**9.66E-10**	4.81E-05	5.94E-05	5.10E-05	4.90E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.38E-05	0.283859	0.10733	-0.0055169	0.0885941	0.0442513
(2)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	117.478	**1.02E-09**	4.81E-05	5.87E-05	5.10E-05	4.25E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.47E-05	0.321145	0.0632796	0.0151823	0.0997437	0.00968388
(3)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	122.966	**1.0986E-09**	4.81E-05	6.35E-05	5.10E-05	4.85E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.82E-05	0.221173	0.10493	0.0344397	0.049225	0.0417724
(4)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	74.8889	**1.1050E-09**	4.81E-05	6.25E-05	5.10E-05	4.48E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.99E-05	0.212318	0.0955053	0.0245537	0.0530928	0.0510944
(5)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	55.8889	**1.13E-09**	4.81E-05	6.32E-05	5.10E-05	4.54E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.91E-05	0.214138	0.0991262	0.0267773	0.0520487	0.0472941
(6)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	53.9468	**1.22E-09**	4.81E-05	6.33E-05	5.10E-05	4.52E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.84E-05	0.218807	0.0997764	0.0255501	0.0515927	0.0504963

	(1)	(2)	(3)	(4)	(5)	(6)
MSE	9.66E-10	1.02E-09	1.0986E-09	1.1050E-09	1.13E-09	1.22E-09
HAC_std	2.43E-10	2.59E-10	2.33E-10	2.37E-10	2.20E-10	2.56E-10
t_MSE	3.96744	3.94106	4.72397	4.66092	5.15286	4.74954
	t12	T13	t14	t15	t25	
T	-0.39802	-1.40348	-1.62466	-1.70449	-0.848255	

Joint Chi-squared test: 48.32

Table 3. We report forecasting regressions of integrated variance (estimated using optimally-defined flat-top modified Tukey-Hanning kernels as in Bandi and Russell, 2010) on five lags of realized variance. The regressor (realized variance) is sampled using 6 methods described in the main text. We use Spiders mid-quotes on the NYSE. The sample period is 2002/1 - 2006/3. In all cases, 1,000 observations are employed to estimate the model parameters (a0 through a5) and forecast. The table reports means and standard deviations for both the regressand (true) and the volatility forecasts. It also reports the average number of observations to be skipped (q) according to each sampling rule. T-statistics for the individual MSEs, t-statistics for pair-wise tests of equal MSEs, and a joint Chi-squared test of equal MSEs across sampling methods are reported in the second panel.

LHS = FlatBart RHS = RV 5 lags
2004/1-2006/3

(1)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	171.112	**2.35E-10**	2.82E-05	3.20E-05	1.52E-05	9.51E-06
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	1.49E-05	0.276181	0.0922763	0.0236755	0.0942341	0.0224669
(2)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	178.018	**2.47E-10**	2.82E-05	3.12E-05	1.52E-05	1.11E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	1.28E-05	0.342192	0.0666968	0.00278203	0.126232	-0.002328
(3)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	160.244	**4.11E-10**	2.82E-05	3.81E-05	1.52E-05	1.48E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.10E-05	0.187816	0.0960295	0.0664505	0.0519591	0.021378
(4)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	65.9439	**5.36E-10**	2.82E-05	3.87E-05	1.52E-05	1.88E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	1.84E-05	0.224782	0.0967559	0.0235081	0.0666108	0.0378135
(5)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	71.2857	**5.57E-10**	2.82E-05	3.88E-05	1.52E-05	1.91E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	1.97E-05	0.220002	0.0931147	0.0254188	0.0662948	0.0317596
(6)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	69.6673	**6.04E-10**	2.82E-05	3.91E-05	1.52E-05	2.06E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	1.87E-05	0.225986	0.0959497	0.0224315	0.0649285	0.0383145

	(1)	(2)	(3)	(4)	(5)	(6)
MSE	2.35E-10	2.47E-10	4.11E-10	5.36E-10	5.57E-10	6.04E-10
HAC_std	2.20E-11	2.52E-11	7.90E-11	1.48E-10	1.49E-10	2.04E-10
t_MSE	10.674	9.80144	5.20368	3.6313	3.73351	2.95806
	t12	t13	t14	t15	t25	
T	-1.54609	-2.38177	-2.09481	-2.21819	-2.11737	

Joint Chi-square test: 117.82

Table 4. We report forecasting regressions of integrated variance (estimated using optimally-sampled flat-top Bartlett kernels as in Bandi and Russell, 2010) on five lags of realized variance. The regressor (realized variance) is sampled using 6 methods described in the main text. We use Spiders mid-quotes on the NYSE. The sample period is 2004/1 - 2006/3. In all cases, 1,000 observations are employed to estimate the model parameters (a0 through a5) and forecast. The table reports means and standard deviations for both the regressand (true) and the volatility forecasts. It also reports the average number of observations to be skipped (q) according to each sampling rule. T-statistics for the individual MSEs, t-statistics for pair-wise tests of equal MSEs, and a joint Chi-squared test of equal MSEs across sampling methods are reported in the second panel.

LHS = FlatCubic RHS = RV 5 lags
2004/1-2006/3

(1)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	171.112	**2.31E-10**	2.94E-05	3.36E-05	1.56E-05	9.47E-06
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	1.63E-05	0.285943	0.100449	0.0191892	0.0940962	0.012773
(2)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	178.018	**2.36E-10**	2.94E-05	3.27E-05	1.56E-05	1.11E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	1.42E-05	0.351297	0.0773286	-0.00088397	0.125422	-0.0128982
(3)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	145.897	**4.90E-10**	2.94E-05	4.08E-05	1.56E-05	1.68E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.27E-05	0.188722	0.100167	0.0738194	0.047709	0.0128716
(4)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	65.9439	**5.65E-10**	2.94E-05	4.04E-05	1.56E-05	1.91E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.00E-05	0.232728	0.104053	0.0209948	0.0620747	0.0318149
(5)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	71.2857	**5.89E-10**	2.94E-05	4.05E-05	1.56E-05	1.94E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.13E-05	0.227631	0.10084	0.0228622	0.0618063	0.0257183
(6)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	69.6673	**6.42E-10**	2.94E-05	4.08E-05	1.56E-05	2.11E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.03E-05	0.234069	0.103272	0.0199102	0.0602224	0.0322606

	(1)	(2)	(3)	(4)	(5)	(6)
MSE	2.31E-10	2.36E-10	4.90E-10	5.65E-10	5.89E-10	6.42E-10
HAC_std	2.31E-11	2.56E-11	1.15E-10	1.68E-10	1.69E-10	2.28E-10
t_MSE	9.97592	9.21065	4.25107	3.36381	3.48204	2.81823
	t12	t13	t14	t15	t25	
T	-0.73573	-2.31294	-2.01153	-2.14223	-2.10502	

Joint Chi-squared test: 105.71

Table 5. We report forecasting regressions of integrated variance (estimated using optimally-defined flat-top cubic kernels as in Bandi and Russell, 2010) on five lags of realized variance. The regressor (realized variance) is sampled using 6 methods described in the main text. We use Spiders mid-quotes on the NYSE. The sample period is 2004/1 - 2006/3. In all cases, 1,000 observations are employed to estimate the model parameters (a0 through a5) and forecast. The table reports means and standard deviations for both the regressand (true) and the volatility forecasts. It also reports the average number of observations to be skipped (q) according to each sampling rule. T-statistics for the individual MSEs, t-statistics for pair-wise tests of equal MSEs, and a joint Chi-squared test of equal MSEs across sampling methods are reported in the second panel.

LHS = FlatTukey RHS = RV 5 lags
2004/1-2006/3

(1)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	171.112	**2.34E-10**	2.94E-05	3.30E-05	1.57E-05	9.43E-06
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	1.58E-05	0.279715	0.103455	0.0165643	0.0901693	0.0233784
(2)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	178.018	**2.42E-10**	2.94E-05	3.23E-05	1.57E-05	1.11E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	1.38E-05	0.346032	0.07832	-0.0041256	0.121751	-0.0022715
(3)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	158.505	**4.55E-10**	2.94E-05	3.96E-05	1.57E-05	1.64E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.20E-05	0.189851	0.103888	0.0649971	0.0479956	0.0198157
(4)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	65.9439	**5.50E-10**	2.94E-05	3.99E-05	1.57E-05	1.90E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	1.94E-05	0.227948	0.105575	0.0190842	0.0605393	0.0394834
(5)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	71.2857	**5.76E-10**	2.94E-05	3.99E-05	1.57E-05	1.93E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.07E-05	0.222864	0.103168	0.0205479	0.0605758	0.0326004
(6)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	69.6673	**6.21E-10**	2.94E-05	4.03E-05	1.57E-05	2.09E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	1.97E-05	0.228984	0.105312	0.0172791	0.0590676	0.0400023

	(1)	(2)	(3)	(4)	(5)	(6)
MSE	2.34E-10	2.42E-10	4.55E-10	5.50E-10	5.76E-10	6.21E-10
HAC_std	2.45E-11	2.71E-11	1.06E-10	1.65E-10	1.66E-10	2.21E-10
t_MSE	9.54519	8.91172	4.3104	3.33834	3.46247	2.80679
	t12	t13	t14	t15	t25	
T	-0.97995	-2.21054	-1.94983	-2.08929	-2.03753	

Joint Chi-squared test: 95.5

Table 6. We report forecasting regressions of integrated variance (estimated using optimally-defined flat-top modified Tukey-Hanning kernels as in Bandi and Russell, 2010) on five lags of realized variance. The regressor (realized variance) is sampled using 6 methods described in the main text. We use Spiders mid-quotes on the NYSE. The sample period is 2004/1 - 2006/3. In all cases, 1,000 observations are employed to estimate the model parameters (a0 through a5) and forecast. The table reports means and standard deviations for both the regressand (true) and the volatility forecasts. It also reports the average number of observations to be skipped (q) according to each sampling rule. T-statistics for the individual MSEs, t-statistics for pair-wise tests of equal MSEs, and a joint Chi-squared test of equal MSEs across sampling methods are reported in the second panel.

LHS = FlatBart RHS = FlatBart 5 lags
2002/1-2006/3

(1)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	10.2688	**9.00E-10**	4.67E-05	5.38E-05	5.00E-05	3.61E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	1.94E-05	0.364522	0.141937	0.0872028	0.099541	0.0647708
(2)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	13.3276	**9.92E-10**	4.67E-05	4.99E-05	5.00E-05	3.17E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.28E-05	0.307943	0.111025	0.0200002	0.091199	0.0587348
(3)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	5.49193	**1.03E-09**	4.67E-05	5.37E-05	5.00E-05	3.48E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.27E-05	0.332203	0.125932	0.0373003	0.09843	0.069047

Table 7. We report forecasting regressions of integrated variance (estimated using optimally-defined flat-top Bartlett kernels as in Bandi and Russell, 2010) on five lags of Bartlett kernel estimates. The regressor's autocovariances are chosen using 3 methods described in the main text. We use Spiders mid-quotes on the NYSE. The sample period is 2002/1 - 2006/3. In all cases, 1,000 observations are employed to estimate the model parameters (a0 through a5) and forecast. The table reports means and standard deviations for both the regressand (true) and the volatility forecasts. It also reports the average number of autocovariances selected according to each rule.

LHS = FlatCubic RHS = FlatCubic 5 lags
2002/1-2006/3

(1)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	8.83476	**1.02E-09**	4.85E-05	5.55E-05	5.19E-05	3.67E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.02E-05	0.367511	0.144537	0.0827307	0.096981	0.0580071
(2)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	13.0978	**1.1814E-09**	4.85E-05	5.18E-05	5.19E-05	3.59E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.39E-05	0.308348	0.112293	0.0176941	0.090176	0.045786
(3)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	5.03989	**1.1908E-09**	4.85E-05	5.52E-05	5.19E-05	3.42E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.32E-05	0.34386	0.127008	0.0337753	0.094211	0.0649336

Table 8. We report forecasting regressions of integrated variance (estimated using optimally-defined flat-top cubic kernels as in Bandi and Russell, 2010) on five lags of cubic kernel estimates. The regressor's autocovariances are chosen using 3 methods described in the main text. We use Spiders mid-quotes on the NYSE. The sample period is 2002/1 - 2006/3. In all cases, 1,000 observations are employed to estimate the model parameters (a0 through a5) and forecast. The table reports means and standard deviations for both the regressand (true) and the volatility forecasts. It also reports the average number of autocovariances selected according to each rule.

LHS = FlatTukey RHS = FlatTukey 5 lags
2002/1-2006/3

(1)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	13.7749	**9.10E-10**	4.81E-05	5.49E-05	5.10E-05	3.67E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	1.94E-05	0.372265	0.150015	0.0671072	0.091612	0.0789061
(2)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	19.8585	**1.0753E-09**	4.81E-05	5.09E-05	5.10E-05	3.24E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.34E-05	0.308512	0.116799	0.0170259	0.087551	0.0556742
(3)	Avg. q	MSE	Mean(Vtrue)	Mean(Vfore)	Std(Vtrue)	Std(Vfore)
	7.94207	**1.0908E-09**	4.81E-05	5.45E-05	5.10E-05	3.38E-05
	Avg. a0	Avg. a1	Avg. a2	Avg. a3	Avg. a4	Avg. a5
	2.27E-05	0.341694	0.131405	0.0328646	0.089792	0.0768419

Table 9. We report forecasting regressions of integrated variance (estimated using optimally-sampled flat-top modified Tukey-Hanning kernels as in Bandi and Russell, 2010) on five lags of flat-top modified Tukey-Hanning kernel estimates. The regressor's autocovariances are chosen using 3 methods described in the main text. We employ Spiders mid-quotes on the NYSE. The sample period is 2002/1 - 2006/3. In all cases, 1,000 observations are employed to estimate the model parameters (a0 through a5) and forecast. The table reports means and standard deviations for both the regressand (true) and the volatility forecasts. It also reports the average number of autocovariances selected according to each rule.