Journal of Business & Economic Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ubes20

Realized Volatility Forecasting in the Presence of Time-Varying Noise

Federico M. Bandi [a d] , Jeffrey R. Russell [b] & Chen Yang [c]

[a] Johns Hopkins University , Baltimore , MD , 21218

[b] University of Chicago , Chicago , IL , 60637

[c] Moody's , New York , NY , 10007

[d] Edhec-Risk , Nice , 06202 , France

Accepted author version posted online: 16 May 2013.Published online: 22 Jul 2013.

To cite this article: Federico M. Bandi , Jeffrey R. Russell & Chen Yang (2013) Realized Volatility Forecasting in the Presence of Time-Varying Noise, Journal of Business & Economic Statistics, 31:3, 331-345, DOI: 10.1080/07350015.2013.803866

To link to this article: http://dx.doi.org/10.1080/07350015.2013.803866

Realized Volatility Forecasting in the Presence of Time-Varying Noise

Federico M. Bandi
Johns Hopkins University, Baltimore, MD 21218, and Edhec-Risk, Nice 06202, France (*fbandi1@jhu.edu*)

Jeffrey R. Russell
University of Chicago, Chicago, IL 60637 (*jeffrey.russell@chicagobooth.edu*)

Chen Yang
Moody's, New York, NY 10007 (*michelle.chen.yang@gmail.com*)

Observed high-frequency financial prices can be considered as having two components, a true price and a market microstructure noise perturbation. It is an empirical regularity, coherent with classical market microstructure theories of price determination, that the second moment of market microstructure noise is time-varying. We study the optimal, from a finite-sample forecast mean squared error (MSE) standpoint, frequency selection for realized variance in linear variance forecasting models with time-varying market microstructure noise. We show that the resulting sampling frequencies are generally considerably lower than those that would be optimally chosen when time-variation in the second moment of the noise is unaccounted for. These optimal, lower frequencies have the potential to translate into considerable out-of-sample MSE gains. When forecasting using high-frequency variance estimates, we recommend treating the relevant frequency as a parameter and evaluating it *jointly* with the parameters of the forecasting model. The proposed joint solution is robust to the features of the true price formation mechanism and generally applicable to a variety of forecasting models and high-frequency variance estimators, including those for which the typical choice variable is a smoothing parameter, rather than a frequency.

KEY WORDS: Realized variance; Time-varying market microstructure noise; Volatility forecasting.

1. INTRODUCTION

The choice of sampling frequency, for realized variance, or bandwidth, for high-frequency kernel estimates of variance, should be dictated by finite-sample methods satisfying suitable, statistical or economic, optimality criteria (Bandi and Russell 2006b). The existing finite-sample work has focused on the period-by-period (named *conditional,* in this article) or *unconditional* mean squared error (MSE) optimization of sampling frequency (Bandi and Russell 2003, 2008; Hansen and Lunde 2006, and Oomen 2006), for realized variance-type objects, or smoothing parameter (Bandi and Russell 2011), for high-frequency kernel estimates of variance. Joint consideration of the frequency/bandwidth selection and forecasting problem is important and was pursued by Andersen, Bollerslev, and Meddahi (2011), ABM hereafter, and Ghysels and Sinko (2011), GS henceforth. In the context of theoretical linear regression models with time-invariant noise, these contributions show that the optimal frequency selection problem, for the purpose of R^2 maximization or forecast MSE minimization, reduces, under assumptions, to the minimization of the unconditional variance of the realized variance estimator.

We re-examine the linear forecasting problem in ABM (2011) and GS (2011) under the assumption of time-variation in the second moment of market microstructure noise (for some early evidence, Bandi and Russell 2006a; Oomen 2006). This time-variation, illustrated graphically in Figure 1 for our data, induces time-variation in the bias of the realized variance estimator constructed using high-frequency data. A time-varying bias in realized variance should have implications for forecasting when using realized variance as a predictor. Intuitively, since the bias is not constant, it cannot be absorbed by the regression's intercept. Hence, minimizing the variance of the regressor (i.e., realized variance), which is the R^2-maximizing solution, under the assumption of a constant bias, may lead to excessively high sampling frequencies and, in light of the one-to-one dependence between sampling frequency and bias contaminations, large biases. Since these biases are time-varying, their dispersion may affect the quality of the forecasts. In essence, finite-sample optimality in the choice of sampling frequency, as given by the R^2-maximizing solution if the noise is assumed to be time-varying, is likely to require a lower sampling frequency, so as to reduce time-changing bias contaminations, than previously derived.

We make three contributions. First, in the context of a classical price formation model with market microstructure noise to which we solely add noise time-variation (Section 1), we formalize the intuition mentioned earlier and derive R^2-optimal frequency selection methods for the joint frequency selection and forecasting problem (Theorem 1). In this context, we find lower optimal frequencies than those derived when time-variation in the noise variance is unaccounted for. Interestingly, the new frequency choices are, in general, closer to those that would be obtained from the optimization of the finite-sample unconditional MSE of the regressor (as in Bandi and Russell 2006a,

Journal of Business & Economic Statistics
July 2013, Vol. 31, No. 3
DOI: 10.1080/07350015.2013.803866

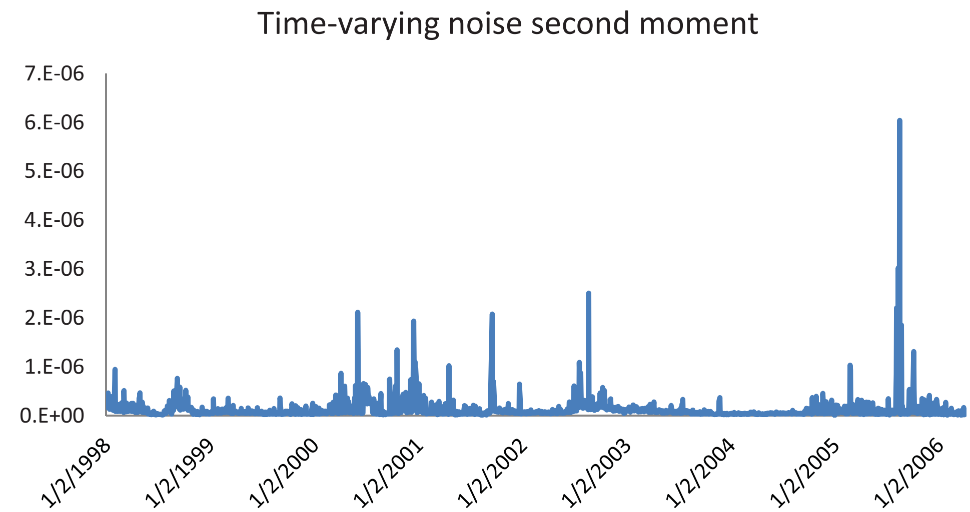

Figure 1. Microstructure noise second moment's estimates. We use Spiders midquotes on the NYSE. The sample period is 1998/1–2006/3. The online version of this figure is in color.

2008) than to those that would be obtained from the optimization of the unconditional variance of the regressor as needed, for forecast MSE minimization, in linear forecasting models under the assumption of a constant noise second moment (ABM, 2011; GS, 2011). In other words, we find that taking bias into account suboptimally (from the point of view of forecasting), through an unconditional MSE-based optimization, again under the assumption of a constant noise second moment, may be an empirically reasonable strategy.

Second, we emphasize that choosing the sampling frequency/bandwidth *conditionally* (for each entry/period in the regressor vector) rather than *unconditionally,* as generally done in the literature, may be a superior strategy from a forecasting standpoint, even when the alternative is the exact, unconditional, R^2-based optimal choice. Again, in the context of a classical price formation mechanism to which we only add noise time-variation, we find empirically verifiable conditions under which this statement is true.

Finally, we recognize that conditional choices may be empirically more cumbersome, in terms of implementation, than unconditional rules. In addition, the conditions under which they would outperform the optimal unconditional choice are, in general, model-specific and, when moving away from our illustrative model to work with more complex specifications, hard to verify in practice. What we propose, instead, for empirical forecasting work using high-frequency data is the *joint* estimation of 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, under the model's assumptions, of course. However, the strategy (i) readily applies to any forecasting model making use of realized variance measures, (ii) is robust to features of the data such as noise dependence, dependence between the noise process and the true price[1] process, and dependence between the true price variance and the noise variance; (iii) does not require estimation of inputs, like the noise second moment variance or the quarticity, for empirical implementation; and (iv) can be broadly applied, by simply changing the choice variable from a frequency to a smoothing parameter, to all of the recently proposed kernel estimates of variance.

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 and Shephard 2002) as well as that of classes of theoretically more robust, to noise, kernel-based estimators (e.g., Zhou 1996; Zhang, Mykland, and Aït-Sahalia 2005; Hansen and Lunde 2006; 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, 2011; Bandi, Russell, and Yang 2008), Mincer-Zarnowitz-style linear regressions (ABM 2004, 2011), or MIDAS-type regressions (GS 2006, 2011). Either statistical metrics, such as forecast MSEs and coefficients of determination (Corradi, Distaso, and Swanson 2011; GS 2006, 2010; Aït-Sahalia and Mancini 2008; ABM 2004, 2011) 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, 2011; Bandi, Russell, and Yang 2008), have been used for the purpose of evaluating the quality of the forecasts. In all of these cases, necessary choice variables when working with high-frequency variance estimates are either the sampling frequency or the degree of smoothing implied by a bandwidth choice. In all cases, both the sampling frequency and the bandwidth can be selected, in agreement with the logic behind the least-squares solution to the linear forecasting problem presented in this article, to directly optimize the statistical or economic criterion of interest.

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

[1]The terminologies "true price" and "equilibrium price" are used interchangeably in this article.

model. Both choices are meant to derive closed-form, conditional and unconditional, results and illustrate, through them, important conceptual issues pertaining to variance forecasting using realized variance measures when the noise variance is time-varying. The theoretical analysis will, however, lead us to our proposed, least-squares, unconditional solution to the choice problem which, as emphasized earlier and discussed in Section 4, 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 FORECASTING MODEL

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

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

or, in terms of continuously compounded returns,

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

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, that is, $p^*_t = \int_0^t \sigma_s dW_s$, 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). 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, when unambiguous, and sometime write V_t instead of $V_{t,t+1}$ for brevity.

We assume the noise contaminations in the price process u are iid in discrete time, over each day, with mean zero, variance σ^2_{tu}, and fourth moment $c_u\sigma^4_{tu}$, where c_u denotes kurtosis. In addition, $u \perp\!\!\!\perp p^*$ and $\sigma^2_{tu} \perp\!\!\!\perp V_t$.[2] 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., Bandi and Russell 2003, 2011; Barndorff-Nielsen et al. 2008; Hansen and Lunde 2006 and Zhang, Mykland, and Aït-Sahalia 2005, among others).[3] While these conditions generally capture important first-order effects in the data, the degree of their empirical accuracy depends on the market structure, on the price measurement (transaction prices vs. midquotes, for example), as well as on the sampling scheme (calendar time vs. event time, for instance). Bandi and Russell (2006b) discussed these ideas.

As said, we depart from the "usual" assumptions by allowing for time-varying, across days, noise moments. This is the sense in which noise is time-varying. Since u captures deviations of observed prices, midquotes, or transaction prices, from equilibrium levels, time-variation in the noise moments is theoretically consistent, as is the case for bid/ask spread determination, with changing degrees of liquidity and asymmetric information (see, e.g., the discussion in Bandi and Russell 2005). Not only is the time-varying nature of the noise moments coherent with classical market microstructure theories of price formation, it is also—barring possible finite-sample contaminations in the corresponding estimates—a widely documented empirical regularity (see Bandi and Russell 2006a and Oomen 2006, for some early evidence).

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 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.[4] We leave the theoretical and empirical complications that this modeling 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 $\widehat{V}_t = \sum_{j=1}^M r^2_{t+j\delta}$ (Barndorff-Nielsen and Shephard 2002; Andersen et al. 2003). To this extent, it is useful to begin with a specific model forming the basis for some of our analysis in the article. 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 on the estimation of the model's parameters.

2.1 Intuition

Under our assumed structure, the realized variance estimator takes the form

$$\widehat{V}_t = \sum_{j=1}^M r^2_{t+j\delta} = \sum_{j=1}^M r^{*2}_{t+j\delta} + \sum_{j=1}^M \varepsilon^2_{t+j\delta} + 2\sum_{j=1}^M r^*_{t+j\delta}\varepsilon_{t+j\delta}.$$

[2]This assumption ($\sigma^2_{tu} \perp\!\!\!\perp V_t$) can be easily relaxed. We will later show how the optimal problem would change if the assumption were not satisfied.

[3]Bandi and Russell (2003, 2008), Aït-Sahalia et al. (2011), Oomen (2005, 2006), and Hansen and Lunde (2006) discuss noise dependence. Kalnina and Linton (2008) allowed for a form of dependence between the noise and the true price.

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

The estimator can also be rewritten as

$$\widehat{V}_t = V_t + M\mathbf{E}\left(\sigma_{t\varepsilon}^2\right) + \left(\sum_{j=1}^{M} r_{t+j\delta}^{*2} - V_t\right) + M\left(\sigma_{t\varepsilon}^2 - \mathbf{E}\left(\sigma_{t\varepsilon}^2\right)\right)$$
$$+ \left(\sum_{j=1}^{M} \varepsilon_{t+j\delta}^2 - M\sigma_{t\varepsilon}^2\right) + 2\sum_{j=1}^{M} r_{t+j\delta}^{*}\varepsilon_{t+j\delta}.$$

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 $\widetilde{a}_t = \sum_{j=1}^{M} \varepsilon_{t+j\delta}^2 - M\sigma_{t\varepsilon}^2$. The unconditional expected bias is given by $M\mathbf{E}(\sigma_{t\varepsilon}^2)$ and the difference between the expected day-t bias and the unconditional expected bias is given by $M(\sigma_{t\varepsilon}^2 - \mathbf{E}(\sigma_{t\varepsilon}^2))$. Finally, we denote the mean-zero, cross-product term by $\gamma_t = 2\sum_{j=1}^{M} r_{t+j\delta}^{*}\varepsilon_{t+j\delta}$.

The forecast error of the estimated model in Equation (1) can now be expressed as

$$\widehat{\xi}_{t+1} = V_{t+1} - (\widehat{\alpha} + \widehat{\beta}\widehat{V}_t) = \left[\alpha - \widehat{\alpha} - \widehat{\beta}M\mathbf{E}\left(\sigma_{t\varepsilon}^2\right)\right]$$
$$+ (\beta - \widehat{\beta})V_t - \widehat{\beta}\left(a_t + \widetilde{a}_t + \gamma_t + M\left(\sigma_{t\varepsilon}^2 - \mathbf{E}\left(\sigma_{t\varepsilon}^2\right)\right)\right)$$
$$+ \xi_{t+1}. \quad (2)$$

For any given value of M, $\widehat{\alpha}$, and $\widehat{\beta}$, the variance of the forecast error is given by

$$\mathrm{var}(\widehat{\xi}_{t+1}) = (\beta - \widehat{\beta})^2 \mathrm{var}(V_t) + \widehat{\beta}^2 \times \left(\underbrace{\underbrace{\underbrace{\mathrm{var}(a_t)}_{\text{No noise}} + \mathrm{var}(\widetilde{a}_t) + \mathrm{var}(\gamma_t)}_{\text{Noise}} + M^2 \mathrm{var}\left(\sigma_{t\varepsilon}^2\right)}_{\text{Time}-\text{varying noise}} \right) + \mathrm{var}(\xi_{t+1}). \quad (3)$$

The sampling frequency only affects the second term in Equation (3) implying that the value of M which minimizes the forecast error variance and the forecast MSE, or maximizes the R^2, can be determined without consideration of the parameters of the forecasting model. Focusing on this second term, we note that the no noise case only yields $\mathrm{var}(a_t)$. This variance is minimized by choosing M as large as possible. The time-invariant noise case leads to $\mathrm{var}(a_t) + \mathrm{var}(\widetilde{a}_t) + \mathrm{var}(\gamma_t)$, a convex function of the sampling frequency (Bandi and Russell 2003, 2008). The solution to this minimization problem is the solution to the joint frequency and forecasting problem in the case of time-invariant noise (ABM, 2011; GS, 2011). The resulting optimal M is lower than the optimal M in the no noise case but, as we show formally in the next section, generally larger than the optimal M in the time-varying noise case due to the presence, in the latter case, of the extra term $M^2\mathrm{var}(\sigma_{t\varepsilon}^2)$. Equation (2), in fact, implies that a time-invariant noise second moment would have no impact on the regressor's variance and would be absorbed by the regression's intercept. It is, however, the variability in $\sigma_{t\varepsilon}^2$ which, when ignored, gives rise to larger-than-optimal M choices and, through the quadratic term $M^2\mathrm{var}(\sigma_{t\varepsilon}^2)$, excessive dispersion of the regressor and the resulting forecast errors.

In sum, *in all cases,* the frequency which minimizes the forecast error variance coincides with the value which minimizes the variance of the regressor ($\widehat{V}_t$). The regressor's variance for the time-varying noise case is, however, our focus.

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

$$\mathrm{var}(\widehat{\xi}_{t+1}) = (\beta - \widehat{\beta})^2 \mathrm{var}(V_t) + \mathrm{var}(\xi_{t+1})$$

and the theoretical solution to the problem would be classical: $\widehat{\beta} = \beta = \mathrm{cov}(V_{t+1}, V_t)/\mathrm{var}(V_t)$. In our case,

$$\widehat{\beta} = \beta \frac{\mathrm{var}(V_t)}{\mathrm{var}(V_t) + \left(\mathrm{var}(a_t) + \mathrm{var}(\widetilde{a}_t) + \mathrm{var}(\gamma_t) + M^2\mathrm{var}(\sigma_{t\varepsilon}^2)\right)} \quad (4)$$

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 $\widehat{\beta} < \beta$.

Next, we consider the R^2-optimal choice problem, for M, in the context of the simple model in this section. Specifically, we provide a discussion in terms of the model's *structural* parameters. Near closed-form expressions will be used to facilitate interpretation before turning to a generally-applicable solution to the frequency choice and forecasting problem (in Section 4).

2.2 Optimal Forecasting Frequencies: Closed-Form Expressions

Theorem 1 presents the optimal, unconditional, rule to choose the R^2—maximizing number of observations M for the proposed, illustrative model with time-varying noise.

Theorem 1. Consider the regression in Equation (1). Under the assumptions made earlier on the data-generating process,

$$M_1 = \arg\max R_M^2 = \arg\min \left\{ \frac{2}{M}\mathbf{E}(Q_t) + \left(2\mathbf{E}(\theta_{t\varepsilon}) - 3\mathbf{E}\left(\sigma_{t\varepsilon}^4\right)\right)M + M^2\mathrm{var}\left(\sigma_{t\varepsilon}^2\right) \right\}, \quad (5)$$

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,[5] this minimization translates into maximization of the forecasting regression's R^2

[5]If the assumption were not satisfied, then the solution would be:

$$M_1^{\blacktriangle} = \arg\max R_M^2 = \arg\max \frac{\left(\mathrm{var}(V_t) + M\mathrm{cov}\left(V_t, \sigma_{t\varepsilon}^2\right)\right)^2}{\mathrm{var}_M(x_t)},$$

where

$$\mathrm{var}_M(x_t) = \frac{2}{M}\mathbf{E}(Q_t) + M\mathbf{E}\left(2\theta_{t\varepsilon} - 3\sigma_{t\varepsilon}^4\right) + \left(4\mathbf{E}\left(\sigma_{t\varepsilon}^2 V_t\right) - \mathbf{E}(\theta_{t\varepsilon})\right.$$
$$\left. + 2\mathbf{E}\left(\sigma_{t\varepsilon}^4\right)\right) + \mathrm{var}(V_t) + M^2\mathrm{var}\left(\sigma_{t\varepsilon}^2\right) + 2M\mathrm{cov}\left(V_t, \sigma_{t\varepsilon}^2\right).$$

Given the proof of Theorem 1, the result is rather obvious.

(as in ABM 2011; GS 2011). 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 Equation (5)).

Remark 2. (Implementation.) The quantities $\theta_{t\varepsilon}$ and $\sigma^2_{t\varepsilon}$ 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).[6] Given $\widehat{\theta}_{t\varepsilon}$ and $\widehat{\sigma}^2_{t\varepsilon}$, consistent estimates of the unconditional moments $\mathbf{E}(\theta_{t\varepsilon})$, $\mathbf{E}(\sigma^4_{t\varepsilon})$, and $\text{var}(\sigma^2_{t\varepsilon})$ can be obtained by employing sample moments of the daily estimates under suitable 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.[7] Roughly unbiased estimates of the unconditional moment $\mathbf{E}(Q_t)$ can then be derived by averaging the estimated daily quarticities under, again, an assumption of stationarity for Q_t. While empirical implementation of the method, by virtue of numerical minimization of the function in Equation (5), is fairly straightforward, the following Corollary provides a convenient, approximate rule to select the optimal M. When we compare it to similar rules in the literature, the new rule will help interpretation, to which we now turn.

Corollary to Theorem 1. For a large optimal M_1,

$$M_1^* \approx \arg\max R^2_M = \left(\frac{\mathbf{E}(Q_t)}{\text{var}(\sigma^2_{t\varepsilon})}\right)^{1/3}. \quad (6)$$

Remark 3. The approximate rule in Equation (6) readily adapts to the noise variance's variance. The larger this variance relative to the signal $\mathbf{E}(Q_t)$ generated by 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. This 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, that is,

$$M_2^* = \left(\frac{\mathbf{E}(Q_t)}{(\mathbf{E}(\varepsilon^2))^2}\right)^{1/3}. \quad (7)$$

It also differs from the optimal (in an R^2 sense), approximate rule proposed by ABM (2011) and GS (2011) in the case of time-invariant noise, that is,[8]

$$M_3^* = \left(\frac{2\mathbf{E}(Q_t)}{2\mathbf{E}(\varepsilon^4) - 3\mathbf{E}(\varepsilon^2)^2}\right)^{1/2}. \quad (8)$$

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

[7]Bandi and Russell (2008) discuss the empirical validity of this simple (albeit theoretically inefficient) procedure by simulation. Efficient estimation of the quarticity is an issue for future work. Important progress on this topic was recently made by Andersen, Dobrev, and Schaumburg (2010).

[8]Unsurprisingly, 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 (Remark 8 in Bandi and Russell 2003 or Remark 4 in Bandi and Russell 2008).

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 $M_2^* = (\frac{\mathbf{E}(Q_t)}{4\sigma_u^4})^{1/3}$ and $M_3^* = (\frac{2\mathbf{E}(Q_t)}{4c_u\sigma_u^4})^{1/2}$ under the above assumptions, but with a time-invariant noise second moment. Intuitively, because the noise-induced bias of the realized variance estimator 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 $\text{var}(\sigma^2_{t\varepsilon}) > (\mathbf{E}(\varepsilon^2))^2 = (\mathbf{E}(\sigma^2_{t\varepsilon}))^2$ 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, in general, $R^2_{M_1^*} > R^2_{M_2^*} > R^2_{M_3^*}$.

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 given by M_3^*. This finding is particularly interesting since the latter choice would in fact be the optimal choice, from a forecast MSE standpoint, should the second moment of the noise, and the realized variance estimator's bias, be assumed to be time-invariant.

While, in this section, we used rules-of-thumb to discuss the main issues, we emphasize that the empirical work, in Section 5, is conducted by optimizing the corresponding full-blown criteria. For example, M_1 in Theorem 1 will be used in place of M_1^* in the Corollary to Theorem 1.

3. *CONDITIONAL* VERSUS *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/period. These period-by-period choices are named *conditional*.[9] Bandi and Russell (2006a, 2008) used this approach, empirically, in predicting variance on the basis of autoregressive, fractionally integrated, models.

This section shows that the conditional approach has the potential to deliver superior forecasts than the unconditional approach described in the previous section. In the context of our illustrative model, whether this is the case depends on empirically verifiable conditions.

We consider the *conditional* finite-sample MSE-based approximate rule in Bandi and Russell (2003, 2008), that is,

$$M_{2t}^* = \left(\frac{Q_t}{\sigma^4_{t\varepsilon}}\right)^{1/3}, \quad (9)$$

and compare it to the approximate optimal unconditional rule in Equation (6).

[9]In this literature, the term conditional is generally short for "conditional on the daily volatility path."

Theorem 2. Define

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

and

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

If

$$\frac{\left(\mathrm{var}(V_t) + \mathrm{cov}\left(V_t, (Q_t)^{1/3}\left(\sigma_{t\varepsilon}^2\right)^{1/3}\right)\right)^2}{\mathrm{var}_{M_{2t}^*}(x_t)} > \frac{(\mathrm{var}(V_t))^2}{\mathrm{var}_{M_1^*}(x_t)}, \tag{10}$$

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 Equation (10)). Leaving aside issues related to estimation uncertainty, the inequality can be easily evaluated empirically by using the methods described in Remark 2 earlier.

Remark 6. Similarly, we can provide a statement for *exact* conditional and unconditional rules. Specifically, one could compare

$$\frac{(\mathrm{var}(V_t))^2}{\mathrm{var}_{M_1}(x_t)} \tag{11}$$

with

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

to

$$\frac{\left(\mathrm{var}(V_t) + \mathrm{cov}\left(V_t, M_{2t}\sigma_{t\varepsilon}^2\right)\right)^2}{\mathrm{var}_{M_{2t}}(x_t)} \tag{12}$$

with

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

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 Equation (12) is larger than Equation (11), 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 the solution of $T+1$, where T is the number of days in the sample, optimization problems to compute the relevant M's (i.e., M_1 and M_{2t} with $t = 1, \ldots, T$). Analogous observations, and derivations, apply to the conditional, approximate minimum variance solution, namely

$$M_{3t}^* = \left(\frac{2Q_t}{2\theta_{t\varepsilon} - 3\sigma_{t\varepsilon}^4}\right)^{1/2}.$$

In both cases, the moment conditions yielding $R^2_{M_{2t}} > R^2_{M_1}$ and $R^2_{M_{3t}} > R^2_{M_1}$ are not satisfied for our data. Consistently, we will show that the in-sample MSEs delivered by conditional choices will, in general, be higher than those yielded by the optimal unconditional choices. However, the out-of-sample MSEs will, in some cases, be lower. We will return to these issues.

4. A GENERAL SOLUTION TO THE UNCONDITIONAL PROBLEM: *JOINT* LEAST-SQUARES OPTIMIZATION

Our previous discussion used a classical price formation mechanism, extended to allow for a time-varying noise second moment, as well as a traditional, albeit simple, forecasting model. Both choices were intended to obtain closed-form implications for important determinants of the optimal frequency choice when noise is time-varying. In this context, we have shown that there are sound theoretical reasons for choosing unconditional sampling frequencies which are lower than those that would be optimally chosen in linear forecasting models when time-variation in the second moment of the noise is unaccounted for. We have also shown that suitable conditional choices have the potential to outperform the R^2-optimal unconditional solution.

While useful for understanding the main issues, the proposed unconditional rule hinges heavily on the assumed, illustrative model. Similarly, the conditions under which conditional choices are to be preferred are, in general, specific to the conditional rule being chosen (M_{2t} or M_{3t}, among other possible choices), the assumed data generating process, and the forecasting model. In all cases, the solutions we derived were intended to minimize the mean squared error of the residuals of the assumed forecasting regressions. We now show that this optimization may be conducted in full generality. Coherently with our previous logic, we now propose to address the optimal, unconditional, frequency/forecasting problem *jointly* with the model's parameters.

To this extent, let $V_{t+1} = f(V_t, V_{t-1}, \ldots, V_{t-(K-1)}|\theta) + \xi_{t+1}$ with $\mathbf{E}(\xi_{t+1}|\mathfrak{F}_t) = 0$ and define $V_{t+1} = f(\widehat{V}_{\widehat{\phi},t}, \widehat{V}_{\widehat{\phi},t-1}, \ldots, \widehat{V}_{\widehat{\phi},t-(K-1)}|\widehat{\theta}) + \widehat{\xi}_{t+1}$, for some function $f(.|\theta)$ and a specific number of lags $K > 0$. We emphasize that the variance estimates $\widehat{V}_{\widehat{\phi},t}$ are a function of ϕ. The parameter ϕ represents the number of observations (or a frequency), and is equal to M, in the case of realized variance. It denotes a bandwidth in the case of high-frequency kernel estimates of variance, like Equation (14) later. The nonlinear least-squares solution to the joint forecasting/sampling problem is given by

$$(\widehat{\phi},\widehat{\theta}) = \arg\min_{\phi,\theta}\sum_{t=1}^{T}\widehat{\xi}_{t+1} = \arg\min_{\phi,\theta}\sum_{t=1}^{T}[V_{t+1}-f(\widehat{V}_{\phi,t},\widehat{V}_{\phi,t-1},\ldots,\widehat{V}_{\phi,t-(K-1)}|\theta)]^2. \tag{13}$$

The approach has several appealing features. First, under the assumptions on the noise in Section 2, if $f(.)$ is consistent with an AR(1) model, the solution to the least-squares problem coincides with the theoretical solution in Theorem 1. More broadly, 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. Second, 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 (2011), to the heterogeneous autoregressive regressions (HAR) of Corsi (2009), possibly with leverage effects and jumps as in Corsi and Renò (2012), and to the HEAVY specifications in Shephard and Sheppard (2009), among other approaches. In the next section, we apply it to an HAR specification. Third, the method does not require the evaluation of objects that are prone to finite-sample bias contaminations, like the second moment of the noise and the quarticity. Finally, it encompasses all available high-frequency variance estimates, those for which the choice variable is a frequency and those for which the choice variable is a smoothing parameter. The issue of feasibility, which has to do with the empirical choice of regressand V_{t+1}, will be discussed in the following applied section.

We note that the specification $f(\widehat{V}_{\phi,t},\widehat{V}_{\phi,t-1},\ldots,\widehat{V}_{\phi,t-(K-1)}|\theta)$ can be viewed as a mixed parameter model in that ϕ is naturally defined over a discrete set. Under conventional assumptions (see, e.g., Ryu 1999 and Choirat and Seri 2012), it is shown that $\mathbf{E}(\widehat{\theta}_i-\theta_{i0})^2 \le \zeta T^{-1}$, for $0 \le i \le$ G and some $\zeta > 0$, and $\mathbf{E}(\widehat{\phi}-\phi_0)^2 \le \rho^T$ for some $0<\rho<1$, where the subscript 0 denotes (pseudo-)true values. 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 our illustrative AR(1) model, the values $\widehat{\theta}=(\widehat{\theta}_2,\widehat{\theta}_1)$ are a slope and an intercept estimate, respectively, consistent for the pseudo-true parameters β_0 in Equation (4) and α_0 (see, Section 2). As shown, β_0 and α_0 do not coincide with the parameters of the true autoregression. For instance, $\beta_0<\beta$. In addition, $\widehat{\phi}$ is consis-

Table 1. **(AR, longer sample)**. We report forecasting regressions of integrated variance (estimated using optimally defined flat-top Bartlett kernels as in Bandi and Russell, 2011) on one lag of realized variance. The regressor (realized variance) is sampled using the six methods described in the main text. We use Spiders midquotes on the NYSE. The sample period is 1/2002–3/2006. In all cases, 1000 observations are employed to estimate the model's parameter and forecast. The table reports the choice of frequency M, the average number of observations to be skipped for each sampling rule q, in-sample and out-of-sample (one-step ahead) MSEs, and in-sample R-squareds. It also reports t-statistics for the individual (in-sample and out-of-sample) MSEs, t-statistics for pairwise tests of equal (in-sample and out-of-sample) MSEs between choice (3) in the main text and all other choices, and a joint chi-squared test of equal (in-sample and out-of-sample) MSEs across sampling methods

	Avg. q	Avg. M	MSE_In	R2_In	MSE_Out	
M_{2t}	162	32	4.66E-09	0.35	1.08E-09	
M_{3t}	117	45	4.52E-09	0.37	1.23E-09	
Min MSE ($\widehat{\phi}=\widehat{M}$)	118	45	4.55E-09	0.37	1.20E-09	
M_1	66	54	4.87E-09	0.32	1.39E-09	
M_2	62	56	4.86E-09	0.32	1.39E-09	
M_3	56	94	4.86E-09	0.32	1.38E-09	
	(1)	(2)	(3)	(4)	(5)	(6)
MSE_In	4.66E-09	4.52E-09	4.55E-09	4.87E-09	4.86E-09	4.86E-09
HAC_std	1.66E-10	1.67E-10	1.63E-10	1.70E-10	1.69E-10	1.69E-10
t_MSE	28.03	27.03	27.88	28.68	28.80	28.72
	t_{13}	t_{23}	t_{43}	t_{53}	t_{63}	
t	18.76	−1.17	28.41	26.39	27.84	
Joint Chi-square test: 2225.18						
	(1)	(2)	(3)	(4)	(5)	(6)
MSE_Out	1.08E-09	1.23E-09	1.20E-09	1.39E-09	1.39E-09	1.38E-09
HAC_std	2.00E-10	3.01E-10	2.15E-10	2.23E-10	2.40E-10	2.11E-10
t_MSE	5.40	4.10	5.59	6.22	5.79	6.57
	t_{13}	t_{23}	t_{43}	t_{53}	t_{63}	
t	−1.34	0.21	2.09	2.75	2.07	
Joint Chi-square test: 101.23						

Table 2. **(AR, shorter sample)**. We report forecasting regressions of integrated variance (estimated using optimally defined flat-top Bartlett kernels as in Bandi and Russell, 2011) on one lag of realized variance. The regressor (realized variance) is sampled using the six methods described in the main text. We use Spiders midquotes on the NYSE. The sample period is 1/2004–3/2006. In all cases, 1000 observations are employed to estimate the model's parameter and forecast. The table reports the choice of frequency M, the average number of observations to be skipped for each sampling rule q, in-sample and out-of-sample (one-step ahead) MSEs, and in-sample R-squareds. It also reports t-statistics for the individual (in-sample and out-of-sample) MSEs, t-statistics for pairwise tests of equal (in-sample and out-of-sample) MSEs between choice (3) in the main text and all other choices, and a joint chi-squared test of equal (in-sample and out-of-sample) MSEs across sampling methods

	Avg. q	Avg. M	MSE_In	R2_In	MSE_Out	
M_{2t}	244	27	2.91E-09	0.46	3.73E-10	
M_{3t}	178	36	2.75E-09	0.49	3.36E-10	
Min MSE ($\widehat{\phi} = \widehat{M}$)	144	45	2.83E-09	0.48	5.21E-10	
M_1	88	49	3.08E-09	0.43	7.01E-10	
M_2	82	51	3.08E-09	0.42	6.69E-10	
M_3	71	92	3.10E-09	0.42	7.98E-10	
	(1)	(2)	(3)	(4)	(5)	(6)
MSE_In	2.91E-09	2.75E-09	2.83E-09	3.08E-09	3.08E-09	3.10E-09
HAC_std	1.83E-10	1.68E-10	1.82E-10	1.92E-10	1.91E-10	1.93E-10
t_MSE	15.91	16.36	15.54	16.08	16.17	16.09
	t_{13}	t_{23}	t_{43}	t_{53}	t_{63}	
t	18.56	−4.36	18.95	16.02	14.52	
Joint Chi-square test: 1298.04						
	(1)	(2)	(3)	(4)	(5)	(6)
MSE_Out	3.73E-10	3.36E-10	5.21E-10	7.01E-10	6.69E-10	7.98E-10
HAC_std	2.93E-11	2.89E-11	9.34E-11	1.18E-10	9.35E-11	1.53E-10
t_MSE	12.75	11.62	5.58	5.93	7.15	5.20
	t_{13}	t_{23}	t_{43}	t_{53}	t_{63}	
t	−1.68	−2.15	2.41	3.10	3.26	
Joint Chi-square test: 175.92						

tent for the value ϕ_0 which minimizes, over the discrete set, the variance of the estimated regressor, as shown in Theorem 1. The same logic applies to the general specification $f(\widehat{V}_{\phi,t}, \widehat{V}_{\phi,t-1}, \ldots, \widehat{V}_{\phi,t-(K-1)}|\theta)$ 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 dispersion of the regressor matrix.

5. FORECASTING REGRESSIONS IN PRACTICE

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

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

$$\widehat{V}_t^{\text{BNHLS}} = \widehat{\gamma}_0 + \sum_{s=1}^{q} w_s(\widehat{\gamma}_s + \widehat{\gamma}_{-s}), \qquad (14)$$

where $\widehat{\gamma}_s = \sum_{j=1}^{M} r_{t+j\delta} r_{t+(j-s)\delta}$ with $s = -q, \ldots, q$, $w_s = k(\frac{s-1}{q})$, and $k(.)$ is a function on [0, 1] satisfying $k(0) = 0$ 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).[11] Furthermore, they have been shown to perform satisfactorily in practice (Bandi, Russell, and Yang 2008 and Bandi and Russell 2011). Importantly, for each day in the sample, the estimators are unbiased under the assumptions in Section 2. This is a useful property

[10]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.

[11]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. See Barndorff-Nielsen et al. (2008) for further discussions.

Table 3. **(HAR, longer sample)**. We report forecasting regressions of integrated variance (estimated using optimally defined flat-top Bartlett kernels as in Bandi and Russell, 2011) on an HAR structure for realized variance. The regressor (realized variance) is sampled using the six methods described in the main text. We use Spiders midquotes on the NYSE. The sample period is 1/2002–3/2006. In all cases, 1000 observations are employed to estimate the model's parameter and forecast. The table reports the choice of frequency M, the average number of observations to be skipped for each sampling rule q, in-sample and out-of-sample (one-step ahead) MSEs, and in-sample R-squareds. It also reports t-statistics for the individual (in-sample and out-of-sample) MSEs, t-statistics for pairwise tests of equal (in-sample and out-of-sample) MSEs between choice (3) in the main text and all other choices, and a joint chi-squared test of equal (in-sample and out-of-sample) MSEs across sampling methods

	Avg. q	Avg. M	MSE_In	R2_In	MSE_Out	
M_{2t}	162	32	4.29E-09	0.41	9.77E-10	
M_{3t}	117	45	4.21E-09	0.42	9.85E-10	
Min MSE ($\widehat{\phi} = \widehat{M}$)	146	36	4.14E-09	0.43	1.11E-09	
M_1	66	54	4.33E-09	0.40	1.06E-09	
M_2	62	56	4.33E-09	0.40	1.05E-09	
M_3	56	94	4.33E-09	0.40	1.09E-09	
	(1)	(2)	(3)	(4)	(5)	(6)
MSE_In	4.29E-09	4.21E-09	4.14E-09	4.33E-09	4.33E-09	4.33E-09
HAC_std	1.54E-10	1.53E-10	1.50E-10	1.54E-10	1.54E-10	1.54E-10
t_MSE	27.82	27.54	27.53	28.10	28.19	28.20
	t_{13}	t_{23}	t_{43}	t_{53}	t_{63}	
t	26.84	5.93	30.70	29.09	32.51	
Joint Chi-square test: 4518.82						
	(1)	(2)	(3)	(4)	(5)	(6)
MSE_Out	9.77E-10	9.85E-10	1.11E-09	1.06E-09	1.05E-09	1.09E-09
HAC_std	2.38E-10	2.36E-10	2.67E-10	2.27E-10	2.29E-10	2.20E-10
t_MSE	4.10	4.18	4.15	4.66	4.56	4.96
	t_{13}	t_{23}	t_{43}	t_{53}	t_{63}	
t	−1.94	−0.87	−0.54	−0.73	−0.16	
Joint Chi-square test: 31.22						

in that it guarantees, theoretically, at least, unbiasedness of the forecasts.[12]

In what follows, we use a Bartlett kernel and optimize the performance of the resulting estimates by using methods discussed in Bandi and Russell (2011). Specifically, for each day in the sample (i.e., *conditionally*, using our previous terminology), we select the number of autocovariances q (or, equivalently, the smoothing sequence) to minimize the estimators' finite-sample variance.[13]

We run regressions of $\widehat{V}_{t+1}^{\text{BNHLS}}$ on lagged realized variance. We consider two forecasting models. The first model is consistent with our illustrative example and simply regresses $\widehat{V}_{t+1}^{\text{BNHLS}}$ on one past value of realized variance:

$$\widehat{V}_{t,t+1}^{\text{BNHLS}} = \widehat{\theta}_1 + \widehat{\theta}_2 \widehat{V}_{M,t-1,t} + \widehat{\xi}_{t+1}.$$

To capture more thoroughly the persistence properties of variance, we also run HAR regressions as in Corsi (2009):

$$\widehat{V}_{t,t+1}^{\text{BNHLS}} = \widehat{\theta}_1 + \widehat{\theta}_2 \widehat{V}_{M,t-1,t} + \widehat{\theta}_3 \widehat{V}_{M,t-5,t} + \widehat{\theta}_4 \widehat{V}_{M,t-22,t} + \widehat{\xi}_{t+1},$$

where $\widehat{V}_{M,t-k,t} = \frac{1}{k}\sum_{i=1}^{k} \widehat{V}_{M,t-i,t-i+1}$ for $k \geq 1$.

We consider two subsets of the data, the full period 1/2002–3/2006 and the shorter 1/2004–3/2006 period.[14] In all cases, we use 1000 observations to estimate the model's parameters and forecast (hence, our data starts in January 1998). We present in-sample and out-of-sample (one-step ahead) forecast MSEs along with related tests (zero MSE values and MSE equality between alternative regressors).

We report six choices of M, leading to six different regressors. The first two are conditional, the remaining four are unconditional. Even though rules-of-thumb are available for all optimal problems with the exception of our more general joint solution in Equation (3), the implementation is conducted, for the sake of superior accuracy, using the corresponding full-blown minimum problems. Coherently, as in the case of M_1 versus its

[12]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.

[13]Other estimators, such as the two-scale estimator of Zhang, Mykland, and Aït-Sahalia (2005) and the multiscale estimator of Zhang (2006), also have favorable properties and could be used.

[14]Results for the 1/2003–3/2006 period are qualitatively identical to those for the 1/2004–3/2006 period and are not provided for brevity.

approximate solution M_1^*, we do not use asterisks to define the M choices next.

1. M_{2t} is the *conditional* solution to the minimum MSE problem in Theorem 2 of Bandi and Russell (2008)
2. M_{3t} is the *conditional* solution to the minimum variance problem in Theorem 4 of Bandi and Russell (2008)
3. $\widehat{\phi} = \widehat{M}$ is the *unconditional* least-squares solution to the maximum R^2 problem or, equivalently, to the minimum forecast MSE problem in Equation (13)
4. M_1 is the *unconditional* solution to the minimum variance problem or, equivalently, the unconditional solution to the maximum R^2 problem when the noise is time-varying in Theorem 1
5. M_2 is the *unconditional* solution to the minimum MSE problem under time-invariant noise, that is, the unconditional version of Theorem 2 in Bandi and Russell (2008)
6. M_3 is the *unconditional* solution to the minimum variance problem or, equivalently, the unconditional solution to the maximum R^2 problem when the noise is time-invariant in ABM (2011) and GS (2011)

Tables 1–4 present results for the AR(1) model and the HAR model, respectively. We begin with the in-sample results. First, we focus on the *unconditional* choices (3–6). For both forecasting models and sample periods, we find $\widehat{\phi} = \widehat{M} < M_1 < M_2 < M_3$. In other words, $q_3 < q_2 < q_1 < \widehat{q}$, where q indicates the average number of high-frequency observations to be skipped in the computation of the corresponding realized variance estimator. Since, empirically, $\text{var}(\sigma^2_{t\varepsilon})(= 5.19e - 14) > (\mathbf{E}(\sigma^2_{t\varepsilon}))^2(= 1.6e - 14)$, the ranking of M values is consistent with the theoretical implications of the rules-of-thumb presented in Section 2. In essence, the variance of the noise second moment leads to variability in the forecast errors which can only be offset by a relatively small choice of M, as is the case for $\widehat{\phi} = \widehat{M}$ and M_1, so as to reduce the impact on the accuracy of the forecasts of a time-changing realized variance's bias.

By construction, the one-step choice 3 must, of course, yield a smaller in-sample MSE than those delivered by choices 4–6. Given a time-varying noise, one would not be surprised if this solution were not very dissimilar, in terms of corresponding MSE values, from the exact theoretical solution M_1 in Equation (4). In fact, under our illustrative model in Section 2, the two choices would exactly coincide. For our data, the in-sample MSE values implied by Equations (4)–(6) are rather similar in spite of the variability of the corresponding sampling frequencies as implied by the alternative M choices. This outcome is evidence

Table 4. **(HAR, shorter sample)**. We report forecasting regressions of integrated variance (estimated using optimally defined flat-top Bartlett kernels as in Bandi and Russell, 2011) on a HAR structure for realized variance. The regressor (realized variance) is sampled using the six methods described in the main text. We use Spiders midquotes on the NYSE. The sample period is 1/2004–3/2006. In all cases, 1000 observations are employed to estimate the model's parameter and forecast. The table reports the choice of frequency M, the average number of observations to be skipped for each sampling rule q, in-sample and out-of-sample (one-step ahead) MSEs, and in-sample R-squareds. It also reports t-statistics for the individual (in-sample and out-of-sample) MSEs, t-statistics for pairwise tests of equal (in-sample and out-of-sample) MSEs between choice (3) in the main text and all other choices, and a joint chi-squared test of equal (in-sample and out-of-sample) MSEs across sampling methods

	Avg. q	Avg. M	MSE_In	R2_In	MSE_Out	
M_{2t}	244	27	2.64E-09	0.51	2.05E-10	
M_{3t}	178	36	2.57E-09	0.52	2.22E-10	
Min MSE ($\widehat{\phi} = \widehat{M}$)	176	37	2.53E-09	0.54	2.30E-10	
M_1	88	49	2.68E-09	0.51	3.36E-10	
M_2	82	51	2.69E-09	0.50	3.12E-10	
M_3	71	92	2.69E-09	0.50	4.03E-10	
	(1)	(2)	(3)	(4)	(5)	(6)
MSE_In	2.64E-09	2.57E-09	2.53E-09	2.68E-09	2.69E-09	2.69E-09
HAC_std	1.69E-10	1.62E-10	1.67E-10	1.70E-10	1.70E-10	1.71E-10
t_MSE	15.60	15.92	15.14	15.71	15.76	15.77
	t_{13}	t_{23}	t_{43}	t_{53}	t_{63}	
t	34.27	3.97	26.26	20.61	19.66	
Joint Chi-square test: 4501.18						
	(1)	(2)	(3)	(4)	(5)	(6)
MSE_Out	2.05E-10	2.22E-10	2.30E-10	3.36E-10	3.12E-10	4.03E-10
HAC_std	2.34E-11	2.54E-11	2.79E-11	7.66E-11	5.90E-11	9.98E-11
t_MSE	8.74	8.74	8.25	4.39	5.30	4.04
	t_{13}	t_{23}	t_{43}	t_{53}	t_{63}	
t	−1.73	−0.55	1.56	1.67	1.93	
Joint Chi-square test: 91.73						

Table 5. **(Simulations, scenario (i))**. We report forecasting regressions of integrated variance (estimated using optimally defined flat-top Bartlett kernels as in Bandi and Russell, 2011) on one lag of realized variance. The regressor (realized variance) is sampled using the six methods described in the main text. We simulate second-by-second prices over 6.5 hr a day for 10 years using the model in the main text. One thousand observations are employed to estimate the model's parameter and forecast. The table reports the choice of frequency M, the average number of observations to be skipped for each sampling rule q, in-sample and out-of-sample (one-step ahead) MSEs, and in-sample R-squareds. It also reports t-statistics for the individual (in-sample and out-of-sample) MSEs, t-statistics for pairwise tests of equal (in-sample and out-of-sample) MSEs between choice (3) in the main text and all other choices, and a joint chi-squared test of equal (in-sample and out-of-sample) MSEs

	Avg. q	Avg. M	MSE_In	R2_In	MSE_Out	
M_{2t}	515	45	5.11E-09	0.82	1.29E-09	
M_{3t}	137	171	5.93E-09	0.82	1.05E-09	
Min MSE ($\widehat{\phi} = \widehat{M}$)	846	28	4.07E-09	0.85	1.37E-09	
M_1	639	37	6.38E-09	0.78	1.67E-09	
M_2	527	45	6.94E-09	0.76	1.85E-09	
M_3	272	86	1.01E-08	0.65	2.73E-09	
	(1)	(2)	(3)	(4)	(5)	(6)
MSE_In	5.11E-09	5.93E-09	4.07E-09	6.38E-09	6.94E-09	1.01E-08
HAC_std	2.05E-10	2.55E-10	1.58E-10	2.58E-10	2.82E-10	4.09E-10
t_MSE	24.92	23.21	25.71	24.71	24.63	24.62
	t_{13}	t_{23}	t_{43}	t_{53}	t_{63}	
t	21.95	19.01	22.82	22.78	23.86	
Joint Chi-square test: 4601.17						
	(1)	(2)	(3)	(4)	(5)	(6)
MSE_Out	1.29E-09	1.05E-09	1.37E-09	1.67E-09	1.85E-09	2.73E-09
HAC_std	1.94E-10	1.54E-10	1.90E-10	2.45E-10	2.88E-10	3.56E-10
t_MSE	6.66	6.82	7.21	6.84	6.42	7.67
	t_{13}	t_{23}	t_{43}	t_{53}	t_{63}	
t	−0.71	−3.04	2.20	2.45	6.66	
Joint Chi-square test: 60.42						

of flatness in the in-sample MSE function across some, less than optimal, frequencies.

Turning to *conditional* choices, we find that using conditional MSE-optimal choices as in Equation (1) and, in some instances, conditional variance-optimal choices as in (2) does not perform in-sample as well as selecting only one frequency for all realized variance entries. As stressed earlier, the theoretical moment condition in Remark 6 is not satisfied for our data.

Statistically, for both forecasting models and sample periods, we find that (i) the in-sample MSEs are always different from zero, (ii) a chi-squared test of the null hypothesis of equal in-sample MSEs across sampling choices is easily rejected, and (iii) pairwise t-tests of the null of equal in-sample MSEs between the least-squares unconditional choice and other unconditional choices reject strongly and consistently.

We now turn to the out-of-sample (one-step ahead) results. First, for both forecasting models and sample periods again, the least-squares solution is generally preferable to the alternative unconditional solutions. This is especially true if the comparison is conducted with respect to M_3, the minimum variance solution under the assumption of a time-invariant noise, in Equation (6) (c.f., t_{63}). In the only instance in which the comparison is found to be favorable to M_3, that is, the HAR model over the longer 2002–2006 period, the statistical difference between the resulting out-of-sample MSEs is, however, determined to be very insignificant ($t_{63} = -0.16$ in Table 3). Second, the conditional solutions (1) and (2) are sometimes found to have favorable out-of-sample MSE properties. Both sets of findings, along with additional insights about the in-sample case, are discussed further by simulation in the next section.

6. MONTE CARLO ANALYSIS

We simulate the same equilibrium price process as in Huang and Tauchen (2005) and, more recently, Nolte and Voev (2012) but without jumps in the price process:

$$dp_t^* = \exp(\beta_0 + \beta_1 v_t) dW_t^{p^*} \qquad p_0^* = 0,$$
$$dv_t = \alpha_v v_t dt + dW_t^v \qquad v_0 = 1/(-2\alpha_v),$$

where $W_t^{p^*}$ and W_t^v are, in agreement with the standard assumptions in Section 2, independent Brownian motions. As is typical in these problems, while empirically warranted, correlating the Brownian shocks does not affect our results in any meaningful way. The parameter vector is

$$\beta_1 = 0.125, \ \beta_0 = \frac{\beta_1^2}{2\alpha_v}, \ \alpha_v = -0.025.$$

Table 6. **(Simulations, scenario (ii))**. We report forecasting regressions of integrated variance (estimated using optimally defined flat-top Bartlett kernels as in Bandi and Russell, 2011) on one lag of realized variance. The regressor (realized variance) is sampled using the six methods described in the main text. We simulate second-by-second prices over 6.5 hr a day for 10 years using the model in the main text. One thousand observations are employed to estimate the model's parameter and forecast. The table reports the choice of frequency M, the average number of observations to be skipped for each sampling rule q, in-sample and out-of-sample (one-step ahead) MSEs, and in-sample R-squareds. It also reports t-statistics for the individual (in-sample and out-of-sample) MSEs, t-statistics for pairwise tests of equal (in-sample and out-of-sample) MSEs between choice (3) in the main text and all other choices, and a joint chi-squared test of equal (in-sample and out-of-sample) MSEs

	Avg. q	Avg. M	MSE_In	R2_In	MSE_Out	
M_{2t}	815	29	2.74E-09	0.88	1.05E-09	
M_{3t}	258	91	4.17E-09	0.85	1.17E-09	
Min MSE ($\widehat{\phi} = \widehat{M}$)	815	29	3.80E-09	0.84	1.88E-09	
M_1	744	35	4.56E-09	0.80	2.22E-09	
M_2	603	43	4.75E-09	0.79	2.51E-09	
M_3	200	117	1.41E-08	0.51	9.41E-09	
	(1)	(2)	(3)	(4)	(5)	(6)
MSE_In	2.74E-09	4.17E-09	3.80E-09	4.56E-09	4.75E-09	1.41E-08
HAC_std	9.79E-11	1.67E-10	1.38E-10	1.65E-10	1.70E-10	5.78E-10
t_MSE	27.98	24.94	27.60	27.58	27.89	24.45
	t_{13}	t_{23}	t_{43}	t_{53}	t_{63}	
t	−26.55	12.18	23.44	26.47	23.45	
Joint Chi-square test: 6152.6						
	(1)	(2)	(3)	(4)	(5)	(6)
MSE_Out	1.05E-09	1.17E-09	1.88E-09	2.22E-09	2.51E-09	9.41E-09
HAC_std	1.35E-10	1.43E-10	1.62E-10	1.90E-10	1.83E-10	5.14E-10
t_MSE	7.74	8.19	11.66	11.66	13.68	18.30
	t_{13}	t_{23}	t_{43}	t_{53}	t_{63}	
t	−9.34	−8.36	3.30	6.03	17.22	
Joint Chi-square test: 355.59						

The noise process u_t is iid $N(0, \sigma_{tu}^2)$ within each day. The variability of the noise second moment is generated in two ways. First, we define the noise second moment in terms of the multiplication of empirically-reasonable, given our data and existing evidence, noise-to-signal ratios and the daily integrated variance of the true price process. Specifically, $\sigma_{tu}^2 = \theta_t V_t$ with $\theta_t = 0.01$ (high ratio) and $\theta_t = 0.0001$ (low ratio). Second, we do the same but replace V_t with its expected value $\mathbf{E}[V_t]$, which is equal to 1 under the model. In both cases, the high-ratio state ($\theta_t = 0.01$) is generated with likelihood 1/3. As emphasized, the noise second moment only changes from day to day and not from observation to observation within a day.

We sample the price process every second over 6.5 hr per day, 252 days per year, and over 10 years. The forecasting model is a simple autoregression. To summarize, we report two scenarios:[15]

(i) In the first scenario, $\sigma_{tu}^2 = \theta_t V_t$. Since the dependence between σ_{tu}^2 and V_t is now apparent, we substitute $M_1^{\blacktriangle}$ for M_1 (as in footnote 5) as our frequency choice (4) to take this dependence into account.[16]

(ii) In the second scenario, $\sigma_{tu}^2 = \theta_t$. Our choice (4) continues to be M_1, that is, the unconditional solution to the maximum R^2 problem when the noise is time-varying (as derived in Theorem 1).

Findings are presented in Table 5 (for scenario (i)) and Table 6 (for scenario (ii)). As earlier, we begin with the in-sample results. In all cases, the least-squares optimal choice (3) leads to drastically lower sampling frequencies, and hence a smaller number of return observations, then the optimal choice under the assumption of a time-invariant noise in Equation (6). The least-squares optimal choice is also generally closer to M_1 or $M_1^{\blacktriangle}$ than to M_2 and M_3. As with data, we find $M_1(M_1^{\blacktriangle}) < M_2 < M_3$ and, in terms of the corresponding

[15]We provide here only a concise set of results. Among other things, we have (i) modified the model parameters (we have, for instance, allowed for correlation between the Brownian shocks), (ii) we have varied the noise variance by changing the probabilities as well as the size of the noise-to-signal ratios, (iii) we have sampled the process at lower 5-sec and 10-sec frequencies, and (iv) we have changed the forecasting model. In all cases, the findings were qualitatively unchanged. Results can be provided upon request.

[16]Importantly, the least-squares solution in Section 4 is immediately robust to this dependence.

in-sample MSE values, $\mathrm{MSE}_{M_3} > \mathrm{MSE}_{M_2} > \mathrm{MSE}_{M_1}(\mathrm{MSE}_{M_1^\blacktriangle})$ with, of course, $\mathrm{MSE}_{M_1}(\mathrm{MSE}_{M_1^\blacktriangle}) > \mathrm{MSE}_{\widehat{M}}$. The ordering of the in-sample MSEs associated with unconditional choices is, as expected, more granular and clearer than with data.

As earlier with data, the conditional choices in Equations (1) and (2) have a tendency to yield larger in-sample MSE values than the optimal unconditional choice in Equation (3), but this is not always the case.

Statistically, we show, again, that (i) the in-sample MSEs are always different from zero, (ii) a chi-squared test of the null hypothesis of equal in-sample MSEs across different sampling choices is easily rejected, and (iii) pairwise t-tests of the null of equal in-sample MSEs between the least-squares unconditional choice and other unconditional choices reject consistently. They do so more strongly when the alternative is choice (6).

Turning now to the out-of-sample outcomes, we find that the performance of the optimal least-squares choice is, as compared to the optimal unconditional choice under the assumption of a time-invariant noise, drastically superior both numerically and statistically (c.f., $t_{63} = 6.66$ in scenario (i) and $t_{63} = 17.22$ in scenario (ii), where, as earlier, the t_{jh}'s are t-statistics for pairwise tests of equal MSEs across method j and method h). As expected, the least-squares optimal MSE choice yields numerical results that are closer to those delivered by the remaining unconditional choices. Even in these cases, however, the numerical differences are in favor of the least-squares optimal MSE choice and are statistically significant. Comparing $\widehat{M}$ to $M_1^\blacktriangle$ and M_1, for instance, we find $t_{43} = 2.2$ in scenario (i) and $t_{43} = 3.3$ in scenario (ii). As with data, the conditional choices in Equations (1) and (2) outperform all unconditional choices, including the least-squares optimal choice.

7. CONCLUSIONS

Price formation mechanisms grounded in classical market microstructure theory imply that the second moment of market microstructure noise is time-varying. Empirical evidence readily confirms this implication of theory. We study the impact of noise time-variation on the forecasting of equilibrium price variance using realized variance. In the context of linear variance forecasting models, we find the need for lower sampling frequencies than required when microstructure noise is assumed to be present but its variability is unaccounted for.

The goal of this article is *not* to advocate a specific variance forecasting model. Choosing an optimal lag structure in the relevant forecasting regressions or enlarging the information set to allow for additional predictors are, among other extensions, important issues beyond the scopes of the present article. 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 conceptual aspects of the volatility forecasting problem in the presence of market microstructure noise which we regard as important. We show that accounting for the time-varying nature of the noise moments through an unconditional least-squares solution which jointly selects the frequency/bandwidth, along with the model's parameters, is theoretically R^2-optimal and may be beneficial in practice from an out-of-sample standpoint. This solution is robust to the features of the true price formation mechanism and is generally applicable to any forecasting model and variance estimator, including those for which the relevant choice variable is a bandwidth, rather than a frequency.

While our focus has been on a well-understood statistical criterion, the same logic leading to a joint least-squares solution to the linear forecasting problem can be applied to suitable economic loss functions. The frequency/bandwidth choice problem can, in fact, be broadly cast in terms of the finite-sample optimization, along with the parameters of the forecasting model, of a variety of metrics of interest. This more general issue is better left for future work.

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 $U_t = \mathbf{E}(\sum_{j=1}^M \varepsilon_j^2|\sigma,\sigma_u) = M\sigma_{t\varepsilon}^2$, and $\eta_t = (\sum_{j=1}^M r_j^{*2} - V_t) + \sum_{j=1}^M r_j^* \varepsilon_j + (\sum_{j=1}^M \varepsilon_j^2 - U_t)$. Notice that $\mathbf{E}(\eta_t|\sigma,\sigma_u) = 0$. Now, consider $R^2 = \frac{\mathrm{cov}^2(V_{t+1},x_t)}{\mathrm{var}(V_{t+1})\mathrm{var}(x_t)}$. The (square root of the) numerator can be expressed as $\mathrm{cov}(V_{t+1},x_t) = \mathrm{cov}(\alpha + \beta V_t + \xi_{t+1}, V_t + U_t + \eta_t) = \beta \mathrm{var}(V_t) + \beta \mathrm{cov}(V_t,\eta_t)$ since $\mathrm{cov}(V_t,U_t) = 0$. But,

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

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

$$\begin{aligned}\mathrm{var}(x_t) &= \mathbf{E}(\mathrm{var}(x_t|\sigma,\sigma_u)) + \mathrm{var}(\mathbf{E}(x_t|\sigma,\sigma_u))\\ &= \mathbf{E}\left(\frac{2}{M}Q_t\right) + \mathbf{E}\left(M\left(2\theta_{t\varepsilon} - 3\sigma_{t\varepsilon}^4\right)\right)\\ &\quad + \left(4\mathbf{E}\left(\sigma_{t\varepsilon}^2 V_t\right) - \mathbf{E}(\theta_{t\varepsilon}) + 2\mathbf{E}\left(\sigma_{t\varepsilon}^4\right)\right)\\ &\quad + \mathrm{var}\left(V_t + M\sigma_{t\varepsilon}^2\right)\\ &= \frac{2}{M}\mathbf{E}(Q_t) + M\mathbf{E}\left(2\theta_{t\varepsilon} - 3\sigma_{t\varepsilon}^4\right)\\ &\quad + \left(4\mathbf{E}\left(\sigma_{t\varepsilon}^2 V_t\right) - \mathbf{E}(\theta_{t\varepsilon}) + 2\mathbf{E}\left(\sigma_{t\varepsilon}^4\right)\right)\\ &\quad + \mathrm{var}\left(V_t\right) + M^2\mathrm{var}\left(\sigma_{t\varepsilon}^2\right),\end{aligned} \tag{A.1}$$

since, again, $\mathrm{cov}(V_t,U_t) = 0$. □

Proof of Theorem 2. Consider Equation (A.1). Plugging in the approximate optimal unconditional rule $M_1^* = (\frac{\mathbf{E}(Q_t)}{\mathrm{var}(\sigma_{t\varepsilon}^2)})^{1/3}$, we obtain

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

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

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

Thus, $R^2_{M_{2t}^*} = \frac{\beta^2(\mathrm{var}(V_t)+\mathrm{cov}(V_t,U_t))^2}{\mathrm{var}(V_{t+1})\mathrm{var}_{M_{2t}^*}(x_t)} = \frac{\beta^2(\mathrm{var}(V_t)+\mathrm{cov}(V_t,(\frac{Q_t}{\sigma_{t\varepsilon}^4})^{1/3}\sigma_{t\varepsilon}^2))^2}{\mathrm{var}(V_{t+1})\mathrm{var}_{M_{2t}^*}(x_t)} = \frac{\beta^2(\mathrm{var}(V_t)+\mathrm{cov}(V_t,(Q_t)^{1/3}(\sigma_{t\varepsilon}^2)^{1/3}))^2}{\mathrm{var}(V_{t+1})\mathrm{var}_{M_{2t}^*}(x_t)}$. □

ACKNOWLEDGMENTS

We thank Torben Andersen, Nour Meddahi, two anonymous referees, the Co-Editor, Jonathan Wright, 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.

[Received February 2011. Revised September 2012.]

REFERENCES

Aït-Sahalia, Y., and Mancini, L. (2008), "Out-of-Sample Forecasts of Quadratic Variation," *Journal of Econometrics*, 147, 17–33. [332]

Aït-Sahalia, Y., Mykland, P., and Zhang, L. (2011), "Ultra High-Frequency Volatility Estimation With Dependent Microstructure Noise," *Journal of Econometrics*, 160, 160–175. [333]

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003), "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625. [332,333]

Andersen, T. G., Bollerslev, T., and Meddahi, N. (2004), "Analytic Evaluation of Volatility Forecasts," *International Economic Review*, 45, 1079–1110. [332]

——— (2011), "Realized Volatility Forecasting and Market Microstructure Noise," *Journal of Econometrics*, 160, 220–234. [331,334,335,340]

Andersen, T. G., Dobrev, D., and Schaumburg, E. (2010), "Integrated Quarticity Estimation: Theory and Practical Implementation," Working paper, Northwestern University, Federal Reserve Board of Governors and Federal Reserve Bank of New York. [335]

Bandi, F. M., and Russell, J. R. (2003), "Microstructure Noise, Realized Variance, and Optimal Sampling," Working paper, University of Chicago. [331,333,334,335,336,344]

——— (2005), "Full-Information Transaction Costs," Working paper, University of Chicago. [333]

——— (2006a), "Separating Microstructure Noise From Volatility," *Journal of Financial Economics*, 79, 655–692. [331,333,335]

——— (2006b), "Comment on Hansen and Lunde," *Journal of Business and Economic Statistics*, 24, 167–173. [331,333]

——— (2007), "Volatility," in *Handbook of Financial Engineering*, eds. J., Birge and V. Linetski, North Holland: Elsevier. [335]

——— (2008), "Microstructure Noise, Realized Variance, and Optimal Sampling," *Review of Economic Studies*, 75, 339–369. [331,332,333,334,335,336,340,343,344]

——— (2011), "Market Microstructure Noise, Integrated Variance Estimators, and the Accuracy of Asymptotic Approximations," *Journal of Econometrics*, 160, 145–159. [331,332,333,338,339]

Bandi, F. M., Russell, J. R., and Yang, C. (2008), "Realized Volatility Forecasting and Option Pricing," *Journal of Econometrics*, 147, 34–46. [332,338]

Barndorff-Nielsen, O. E., Hansen, P., Lunde, A., and Shephard, N. (2008), "Designing Realized Kernels to Measure Ex-Post Variation of Equity Prices in the Presence of Noise," *Econometrica*, 76, 1481–1536. [332,333,338]

Barndorff-Nielsen, O. E., and 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. [332,333]

Choirat, C., and Seri, R. (2012), "Estimation in Discrete Parameter Models," *Statistical Science*, 27, 278–293. [337]

Corradi, V., Distaso, W., and Swanson, N. (2011), "Predictive Inference for Integrated Volatility," *Journal of the American Statistical Association*, 106, 1496–1512. [332]

Corsi, F. (2009), "A Simple Approximate Long-Memory Model of Realized Volatility," *Journal of Financial Econometrics*, 7, 174–196. [337,339]

Corsi, F., and Renò, R. (2012), "Discrete-Time Volatility Forecasting With Persistent Leverage Effect and the Link With Continuous-Time Volatility Modeling," *Journal of Business and Economic Statistics*, 30, 368–380. [337]

Ghysels, E., and Sinko, A. (2006), "Comment on Hansen and Lunde," *Journal of Business and Economic Statistics*, 24, 192–194. [332]

——— (2011), "Volatility Forecasting and Microstructure Noise," *Journal of Econometrics*, 160, 257–271. [331,332,334,335,337,340]

Hansen, P. R., and Lunde, A. (2006), "Realized Variance and Market Microstructure Noise" (with discussions), *Journal of Business and Economic Statistics*, 24, 127–161. [331,332,333]

Huang, X., and Tauchen, G. (2005), "The Relative Contribution of Jumps to Total Price Variance," *Journal of Financial Econometrics*, 3, 456–499. [341]

Kalnina, I., and Linton, O. (2008), "Estimating Quadratic Variation Consistently in the Presence of Correlated Measurement Error," *Journal of Econometrics*, 147, 47–59. [333]

Nolte, I., and Voev, V. (2012), "Least-Squares Inference on Integrated Volatility and the Relationship Between Efficient Prices and Noise," *Journal of Business and Economic Statistics*, 30, 94–108. [341]

Oomen, R. C. A. (2005), "Properties of Bias-Corrected Realized Variance Under Alternative Sampling Schemes," *Journal of Financial Econometrics*, 3, 555–577. [331,333]

——— (2006), "Properties of Realized Variance Under Alternative Sampling Schemes," *Journal of Business and Economic Statistics*, 24, 219–237. [331,333]

Ryu, K. (1999), "Econometric Analysis of Mixed Parameter Models," Working Paper. [337]

Shephard, N., and Sheppard, K. (2009), "Realising the Future: Forecasting With High-Frequency Based (HEAVY) Volatility Models," *Journal of Applied Econometrics*, 23, 197–231. [337]

Zhang, L. (2006), "Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach," *Bernoulli*, 12, 1019–1043. [339]

Zhang, L., Mykland, P., and 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. [332,333,339]

Zhou, B. (1996), "High-Frequency Data and Volatility in Foreign-Exchange Rates," *Journal of Business and Economic Statistics*, 14, 45–52. [332]