Realized volatility forecasting and market microstructure noise[☆]

Torben G. Andersen [a,b,c,*], Tim Bollerslev [d,b,c,1], Nour Meddahi [e,2]

[a] *Department of Finance, Kellogg School of Management, Northwestern University, Evanston, IL 60208, United States*

[b] *NBER, Cambridge, MA, United States*

[c] *CREATES, Aarhus, Denmark*

[d] *Department of Economics, Duke University, Durham, NC 27708, United States*

[e] *Toulouse School of Economics (GREMAQ, IDEI), 21 Allée de Brienne, 31000 Toulouse, France*

ARTICLE INFO

ABSTRACT

We extend the analytical results for reduced form realized volatility based forecasting in ABM (2004) to allow for market microstructure frictions in the observed high-frequency returns. Our results build on the eigenfunction representation of the general stochastic volatility class of models developed by Meddahi (2001). In addition to traditional realized volatility measures and the role of the underlying sampling frequencies, we also explore the forecasting performance of several alternative volatility measures designed to mitigate the impact of the microstructure noise. Our analysis is facilitated by a simple unified quadratic form representation for all these estimators. Our results suggest that the detrimental impact of the noise on forecast accuracy can be substantial. Moreover, the linear forecasts based on a simple-to-implement 'average' (or 'subsampled') estimator obtained by averaging standard sparsely sampled realized volatility measures generally perform on par with the best alternative robust measures.

1. Introduction

In recent years, the increased availability of complete transaction and quote records for financial assets has spurred a literature seeking to exploit this information in estimating the current level of return volatility. Merton (1980) notes that spot volatility may be inferred perfectly if the asset price follows a diffusion process and a continuous record of prices is available. However, practical implementation presents significant challenges. First, we only observe prices at intermittent and discrete points in time. This induces discretization errors in estimates of current volatility. Second, and more importantly, the recorded prices do not reflect direct observations of a frictionless diffusive process. Market prices are quoted on a discrete price grid with a gap between buying and selling prices, i.e., a bid-ask spread, and different prices may be quoted simultaneously by competing market makers due to heterogeneous beliefs, information and inventory positions. The latter set of complications is referred to jointly as market microstructure effects. Consequently, any observed price does not represent a unique market price but instead an underlying ideal price confounded by an error term reflecting the impact of market microstructure frictions, or "noise".

The early literature accommodates microstructure noise by sampling prices relatively sparsely to ensure that the intraday returns are approximately mean zero and uncorrelated. As such, the realized volatility estimator, which cumulates intraday squared returns, provides a near unbiased return variation measure; see, e.g., Andersen et al. (2000). Further, as stressed by Andersen and Bollerslev (1998), Andersen et al. (2001), and Barndorff-Nielsen and Shephard (2001), in the diffusive case, and absent microstructure frictions, this estimator is consistent for the integrated variance as the sampling frequency diverges. Importantly, this represents a paradigm shift towards ex-post estimation of (average)

volatility over a non-trivial interval, avoiding the pitfalls associated with estimation of spot volatility when prices embed microstructure distortions.

However, realized volatility computed from sparsely sampled data suffers from a potentially substantial discretization error; see Barndorff-Nielsen and Shephard (2002), Jacod and Protter (1998), and Meddahi (2002a). An entire literature is devoted to improving the estimator. Important contributions include the first-order autocorrelation adjustment by Zhou (1996), the notion of an optimal sampling frequency by Bandi and Russell (2006, 2008) and Aït-Sahalia et al. (2005), the average and two-scale estimator of Zhang et al. (2005), the multi-scale estimator of Zhang (2006), as well as the realized kernel estimator of Barndorff-Nielsen et al. (2008a).

Another key issue concerns the use of realized volatility measures for decision making. Real-time asset allocation, derivatives pricing and risk management is conducted given current (conditional) expectations for the return distribution over the planning horizon. Hence, the measures of current and past volatility must be converted into useful predictors of a future return variation. This critical step is inevitably model-dependent but the realized volatility based forecasting literature is less developed. A number of empirical studies compares the performance of forecasts using realized variation measures to standard stochastic volatility (SV) forecasts as well as option based predictions; see Andersen et al. (2003), Deo et al. (2006), Koopman et al. (2005), and Pong et al. (2004), among others. The realized variation forecasts generally dominate traditional SV model forecasts based on daily data and they perform roughly on par with the options based forecasts. In terms of a more analytic assessment, existing results stem from a handful of simulation studies which, aside from being model specific, typically ignore microstructure effects.[3]

The model-specific nature of these studies is partially circumvented by Andersen, Bollerslev and Meddahi (henceforth ABM, 2004, 2005). They exploit the eigenfunction stochastic volatility (ESV) framework of Meddahi (2001) in developing analytic expressions for forecast performance spanning all SV diffusions commonly used in the literature. This set-up delivers expressions for the optimal linear forecasts based on the history of past realized volatility measures and allows for direct comparison as the sampling frequency of the intraday returns varies or the measurement horizon changes.[4] It also facilitates analysis of the (artificial) deterioration in forecast performance due to the use of feasible realized volatility measures as ex-post benchmarks for return variation in lieu of the true integrated volatility. Nonetheless, these studies do not account for the impact of microstructure noise on practical measurement and forecast performance. In fact, there is no obvious way to assess this issue analytically for a broad class of models within the existing literature.[5]

In this paper, we extend the ABM studies by explicitly accounting for microstructure noise in the analytic derivation of realized volatility based forecasts. The literature on this topic is limited to concurrent work by Aït-Sahalia and Mancini (2008) and Ghysels and Sinko (2006). These papers provide complementary evidence as they resort to simulation methods or empirical assessment in order to rank the estimators while also studying data generating

processes and forecast procedures not considered here.[6] For example, Aït-Sahalia and Mancini (2008) include long memory and jump diffusions among the scenarios explored, while Ghysels and Sinko (2006) consider nonlinear forecasting techniques based on the MIDAS regression approach. Moreover, a preliminary review of some results, originally derived for this project, is included in Garcia and Meddahi (2006).

The remainder of the paper unfolds as follows. The next section briefly introduces the theoretical framework, including the ESV model and the definition of realized volatility, followed by an enumeration of the analytical expressions for the requisite moments underlying our main theoretical results. Section 3 presents the optimal linear forecasting rules for integrated volatility when the standard realized volatility measure is contaminated by market microstructure noise. We also quantify the impact of noise for forecast performance and explore notions of "optimal" sampling frequency. Moreover, we show how optimally combining intraday squared returns in constructing integrated volatility forecasts does not materially improve upon forecasts relying on realized volatilities computed from equally weighted intraday squared returns. Section 4 shows that many robust realized volatility measures, designed to mitigate the impact of microstructure noise, may be conveniently expressed as a quadratic form of intraday returns sampled at the highest possible frequency. This representation, in turn, facilitates the derivation of the corresponding optimal linear integrated volatility forecasts. We find that a simple estimator, obtained by averaging different sparsely sampled standard realized volatility measures (sometimes referred to as a subsampled estimator), is among the best forecast performers. Moreover, the differences among the competing realized volatility estimators can be substantial, highlighting the potential impact of noise for practical forecast performance. For example, we show that feasible realized volatility forecasting regressions based on the "wrong" realized volatility measure may, falsely, suggest near zero predictability, when in fact more than fifty percent of the day-to-day variation in the (latent) integrated volatility is predictable. Section 5 provides concluding remarks. All main proofs are deferred to the technical Appendix.

2. Theoretical framework

2.1. General setup and assumptions

We focus on a single asset traded in a liquid financial market. We assume that the sample-path of the corresponding (latent) price process, $\{S_t^*, 0 \le t\}$, is continuous and determined by the stochastic differential equation (sde)

$$d \log(S_t^*) = \sigma_t dW_t, \tag{2.1}$$

where W_t denotes a standard Brownian motion, and the spot volatility process σ_t is predictable and has a continuous sample path. We assume the σ_t and W_t processes are uncorrelated and, for convenience, we refer to the unit time interval as a day.[7]

Our primary interest centers on forecasting the (latent) integrated volatility over daily and longer inter-daily horizons. Specifically, we define the one-period integrated volatility,

$$\mathrm{IV}_{t+1} \equiv \int_t^{t+1} \sigma_\tau^2 d\tau, \tag{2.2}$$

---

[3] For example, Andersen et al. (1999) document substantial gains from volatility forecasts based on high-frequency data over daily GARCH forecasts through simulations from a GARCH diffusion.

[4] This same approach has recently been adopted by Corradi et al. (2009a,b) in analyzing the predictive inference for integrated volatility.

[5] Bandi et al. (2008) find that choosing a proper sampling frequency in constructing realized volatility measures has important benefits for a dynamic portfolio choice.

[6] We only became aware of these projects after initiating the current work.

[7] The extension to scenarios including either return–volatility correlations or a drift term in the return process is discussed in the working paper version of this paper (ABM, 2006).

and, for m a positive integer, the corresponding multi-period measure,

$$\text{IV}_{t+1:t+m} = \sum_{j=1}^{m} \text{IV}_{t+j}. \tag{2.3}$$

In this context, IV_t equals the quadratic return variation which, in turn, provides a natural measure of the ex-post return variability; see, e.g., the discussion in Andersen et al. (2006, 2010).[8]

Integrated volatility is not directly observable but, as highlighted by Andersen and Bollerslev (1998), Andersen et al. (2001, 2003), Barndorff-Nielsen and Shephard (2001, 2002), and Meddahi (2002a), the corresponding realized volatilities provide consistent estimates of IV_t. The standard realized volatility measure is simply,

$$\text{RV}_t^*(h) \equiv \sum_{i=1}^{1/h} r_{t-1+ih}^{*(h)2}, \tag{2.4}$$

where $1/h$ is assumed to be a positive integer and

$$r_t^{*(h)} \equiv \log(S_t^*) - \log(S_{t-h}^*). \tag{2.5}$$

Formally, $\text{RV}_t^*(h)$ is uniformly consistent for IV_t as $h \rightarrow 0$, i.e., the intraday sampling frequency goes to infinity. Moreover, ABM (2004) demonstrate that simple autoregressive models for $\text{RV}_t^*(h)$ provide simple-to-implement and, for many popular SV models, remarkably close to efficient forecasts for IV_{t+1} and $\text{IV}_{t+1:t+m}$.[9]

In practice, recorded prices are invariably affected by microstructure frictions and do not adhere to the model in (2.1) at the highest frequencies. The studies above advocate using relatively sparse sampling to allow Eq. (2.1) to adequately approximate the observed price process (see, e.g., the discussion of the so-called volatility signature plot in Andersen et al., 2000, for informally selecting the value of h). More recently, however, many studies advocate explicitly including noise terms in the price process and then design procedures to mitigate their impact on volatility measurement.

We focus on the most common scenario in the literature involving *i.i.d.* noise. Hence, the observed price, $\{S_t, 0 \leq t\}$, is governed by the process in (2.1) *plus* a noise component,

$$\log(S_t) = \log(S_t^*) + u_t, \tag{2.6}$$

where u_t is *i.i.d.*, independent of the frictionless price process S_t^*, with mean zero, variance V_u, and kurtosis $K_u = E[u_t^4]/V_u^2$. In the illustrations below we focus on $K_u = 3$, corresponding to a Gaussian noise term, but our results allow for any finite value of K_u.[10]

If S_t, but not S_t^*, is observable, the h-period returns become,

$$r_t^{(h)} \equiv \log(S_t) - \log(S_{t-h}). \tag{2.7}$$

These noise contaminated returns are linked to the returns in (2.5) as,

$$r_t^{(h)} = r_t^{*(h)} + e_t^{(h)}, \tag{2.8}$$

where

$$e_t^{(h)} \equiv u_t - u_{t-h}. \tag{2.9}$$

The noise induces an MA(1) error structure in observed returns. For very small h the variance of the noise term, $e_t^{(h)}$, dominates the variance of the true return, $r_t^{*(h)}$. In fact, as shown by Bandi and Russell (2006, 2008) and Zhang et al. (2005), the feasible realized volatility measure based on contaminated high-frequency returns,

$$\text{RV}_t(h) \equiv \sum_{i=1}^{1/h} r_{t-1+ih}^{(h)2} \tag{2.10}$$

is inconsistent for IV_t and diverges to infinity for $h \rightarrow 0$. Nonetheless, $\text{RV}_t(h)$ can still be used to construct meaningful forecasts for IV_{t+1} and $\text{IV}_{t+1:t+m}$ for moderate values of h. Indeed, as documented below, by balancing the impact of the noise and signal, accurate volatility forecasting based on simple autoregressive models for $\text{RV}_t(h)$ is feasible. In addition, a number of alternative robust realized volatility measures, explicitly designed to account for high-frequency noise, has been proposed. We therefore also compare and contrast the performance of reduced form forecasting models for these alternative measures to those based on the traditional $\text{RV}_t(h)$ measure.

2.2. Eigenfunction stochastic volatility models

We follow ABM (2004) in assuming that the spot volatility process belongs to the Eigenfunction Stochastic Volatility (ESV) class introduced by Meddahi (2001). This class of models includes most diffusive stochastic volatility models in the literature.

For illustration, assume that volatility is driven by a single state variable.[11] The ESV representation, with p denoting a positive, possibly infinite, integer, takes the generic form,

$$\sigma_t^2 = \sum_{n=0}^{p} a_n P_n(f_t), \tag{2.11}$$

where the latent state variable f_t is governed by a diffusion process,

$$df_t = m(f_t)dt + \sqrt{v(f_t)}dW_t^f, \tag{2.12}$$

the Brownian motion, W_t^f, is independent of W_t in Eq. (2.1), the a_n coefficients are real numbers and the $P_n(f_t)$'s are the eigenfunctions of the infinitesimal generator associated with f_t.[12] The eigenfunctions are orthogonal and centered at zero,

$$E\left[P_n(f_t)P_j(f_t)\right] = 0 \qquad E[P_n(f_t)] = 0, \tag{2.13}$$

and follow first-order autoregressive processes,

$$\forall l > 0, \quad E\left[P_n(f_{t+l}) \mid f_\tau, \tau \leq t\right] = \exp(-\lambda_n l)P_n(f_t), \tag{2.14}$$

where $(-\lambda_n)$ denote the corresponding eigenvalues. These simplifying features render a derivation of analytic multi-step forecasts for σ_t^2 and for the moments of discretely sampled returns, $r_t^{(h)}$, from the model defined by Eqs. (2.1), (2.8), (2.11) and (2.12), feasible. The following proposition collects the relevant results for the subsequent analysis.

Proposition 2.1. *Let the discrete-time noise-contaminated and ideal returns, $r_t^{(h)}$ and $r_t^{*(h)}$, respectively, be determined by an ESV model and Eq. (2.8), with corresponding one- and m-period integrated volatilities, IV_t and $\text{IV}_{t+1:t+m}$, defined by Eqs. (2.2) and (2.3).Then for*

---

[8] The integrated volatility also figures prominently in the option pricing literature; see, e.g., Hull and White (1987) and Garcia et al. (2011).

[9] These theoretical results corroborate the empirical findings of Andersen et al. (2003), Areal and Taylor (2002), Corsi (2009), Deo et al. (2006), Koopman et al. (2005), Martens (2002), Pong et al. (2004), and Thomakos and Wang (2003), among many others, involving estimation of reduced form forecasting models for various realized volatility series.

[10] In addition, the case of correlated microstructure noise is also briefly discussed in ABM (2006).

[11] The one-factor ESV model may be extended to allow for multiple factors while maintaining the key results discussed below; see Meddahi (2001) for further details. See also Chen et al. (2009) for a general approach to eigenfunction modeling for multivariate Markov processes.

[12] For a more detailed discussion of the properties of infinitesimal generators see, e.g., Hansen and Scheinkman (1995) and Aït-Sahalia et al. (2010).

positive integers $i \geq j \geq k \geq l$, $m \geq 1$, and $h > 0$,

$$E[r_{t+ih}^{(h)}] = E[r_{t+ih}^{*(h)}] = 0, \tag{2.15}$$

$$E[r_{t+ih}^{(h)2}] = \text{Var}[r_{t+ih}^{(h)}] = \text{Var}[r_{t+ih}^{*(h)}] + 2V_u = a_0 h + 2V_u, \tag{2.16}$$

$$\text{Cov}[r_{t+ih}^{(h)}, r_{t+jh}^{(h)}] = -V_u, \quad \text{for } |i-j| = 1, \tag{2.17}$$

$$E[r_{t+ih}^{(h)} r_{t+jh}^{(h)} r_{t+kh}^{(h)} r_{t+lh}^{(h)}]$$

$$= 3a_0^2 h^2 + 6 \sum_{n=1}^{p} \frac{a_n^2}{\lambda_n^2}[-1 + \lambda_n h + \exp(-\lambda_n h)]$$

$$\quad + 2V_u^2(K_u + 3) + 12a_0 V_u h \quad \text{if } i = j = k = l,$$

$$= -V_u^2(K_u + 3) - 3a_0 V_u h \quad \text{if } i = j = k = l+1$$

$$\quad \text{or } i = j+1 = k+1 = l+1,$$

$$= a_0^2 h^2 + \sum_{n=1}^{p} \frac{a_n^2}{\lambda_n^2}[1 - \exp(-\lambda_n h)]^2 + V_u^2(K_u + 3) + 4a_0 V_u h$$

$$\quad \text{if } i = j = k+1 = l+1,$$

$$= a_0^2 h^2 + \sum_{n=1}^{p} \frac{a_n^2}{\lambda_n^2}[1 - \exp(-\lambda_n h)]^2 \exp(-\lambda_n(i-k-1)h)$$

$$\quad + 4V_u^2 + 4a_0 V_u h \quad \text{if } i = j > k+1, \ k = l,$$

$$= 2V_u^2 \quad \text{if } i = j+1, \ j = k = l+1,$$

$$= -2V_u^2 - a_0 V_u h \quad \text{if } i = j \geq k+1,$$

$$\quad k = l+1, \ \text{or } i = j+1, \ j \geq k+1, \ k = l,$$

$$= V_u^2 \quad \text{if } i = j+1, \ j \geq k+1, \ k = l+1,$$

$$= 0 \quad \text{otherwise.} \tag{2.18}$$

$$\text{Cov}[r_{t-1+m+ih}^{(h)} r_{t-1+m+jh}^{(h)}, r_{t-1+kh}^{(h)} r_{t-1+lh}^{(h)}]$$

$$= \sum_{n=1}^{p} \frac{a_n^2}{\lambda_n^2}(1 - \exp(-\lambda_n h))^2 \exp(-\lambda_n(m + (i-k-1)h))$$

$$\quad \text{if } m \geq 2, \ i = j, \ k = l,$$

$$= \sum_{n=1}^{p} \frac{a_n^2}{\lambda_n^2}(1 - \exp(-\lambda_n h))^2 \exp(-\lambda_n(1 + (i-k-1)h))$$

$$\quad \text{if } m = 1, \ i = j, \ k = l, \ i \neq 1, k \neq 1/h,$$

$$= \sum_{n=1}^{p} \frac{a_n^2}{\lambda_n^2}(1 - \exp(-\lambda_n h))^2 + (K_u - 1)V_u^2$$

$$\quad \text{if } m = 1, i = j, k = l, i = 1, k = 1/h,$$

$$= 0 \quad \text{otherwise.} \tag{2.19}$$

$$\text{Cov}\left[\text{IV}_t, r_{t-1+ih}^{(h)} r_{t-1+jh}^{(h)}\right]$$

$$= \delta_{i,j}\left(2\sum_{n=1}^{p} \frac{a_n^2}{\lambda_n^2}[\exp(-\lambda_n h) + \lambda_n h - 1]\right.$$

$$\quad + \sum_{n=1}^{p} \frac{a_n^2}{\lambda_n^2}[2 - \exp(-\lambda_n(i-1)h)$$

$$\quad \left. - \exp(-\lambda_n(1-ih))][1 - \exp(-\lambda_n h)]\right). \tag{2.20}$$

$$\text{Cov}[\text{IV}_{t+1:t+m}, r_{t-1+ih}^{(h)} r_{t-1+jh}^{(h)}]$$

$$= \delta_{i,j}\left(\sum_{i=1}^{p} \frac{a_i^2}{\lambda_i^2}[1 - \exp(-\lambda_n h)][1 - \exp(-\lambda_n m)]\right.$$

$$\quad \left. \times \exp(-\lambda_n(1-ih))\right). \tag{2.21}$$

For illustration, we rely on a GARCH diffusion and a two-factor affine model which are representative of the literature. We provide their sde form below, while the corresponding ESV specifications are given in ABM (2004, 2005).

Model M1—GARCH diffusion. The instantaneous volatility is given by,

$$d\sigma_t^2 = \kappa(\theta - \sigma_t^2)dt + \sigma\sigma_t^2 dW_t^{(2)}, \quad \text{where } \kappa = 0.035,$$
$$\theta = 0.636, \ \psi = 0.296. \tag{2.22}$$

Model M2—two-factor affine. The instantaneous volatility is given by,

$$\sigma_t^2 = \sigma_{1,t}^2 + \sigma_{2,t}^2 \qquad d\sigma_{j,t}^2 = \kappa_j(\theta_j - \sigma_{j,t}^2)dt + \eta_j\sigma_{j,t}dW_t^{(j+1)}, \tag{2.23}$$
$$j = 1, 2,$$

where $\kappa_1 = 0.5708$, $\theta_1 = 0.3257$, $\eta_1 = 0.2286$, $\kappa_2 = 0.0757$, $\theta_2 = 0.1786$, and $\eta_2 = 0.1096$, implying a very volatile first factor and a much more slowly mean reverting second factor, reminiscent of, e.g., the Engle and Lee (1999) model.

3. Traditional realized volatility based forecasts

3.1. Optimal linear forecasts

ABM (2004) examine integrated volatility forecasts constructed from linear regressions of IV_{t+1} and $\text{IV}_{t+1:t+m}$ on current and lagged values of $\text{RV}_t^*(h)$. However, Eq. (2.6) provides a better description of real-world prices than (2.1) when data are sampled at the highest frequencies. Hence, for small h, analytical results based on $\text{RV}_t(h)$ in lieu of $\text{RV}_t^*(h)$ should help us gauge the impact of microstructure noise on forecast accuracy.

In order to derive analytical expressions for linear forecasts of IV_{t+1} and $\text{IV}_{t+1:t+m}$ based on current and lagged $\text{RV}_t(h)$ we must compute $\text{Cov}[\text{IV}_{t+1}, \text{RV}_{t-l}(h)]$, $\text{Var}[\text{RV}_t(h)]$, and $\text{Cov}[\text{RV}_{t+1}(h), \text{RV}_{t-l}(h)]$, for $l \geq 0$. To do so, note from Eq. (2.8) that,

$$\text{RV}_t(h) = \text{RV}_t^*(h) + \sum_{i=1}^{1/h} e_{t-1+ih}^{(h)2} + 2\sum_{i=1}^{1/h} r_{t-1+ih}^{*(h)} e_{t-1+ih}^{(h)}. \tag{3.1}$$

Utilizing this decomposition the following set of results follows readily.

Proposition 3.1. *Let the discrete-time noise-contaminated and ideal returns, $r_t^{(h)}$ and $r_t^{*(h)}$ be given by Eqs. (2.8) and (2.5), with the corresponding realized and integrated volatilities, $\text{RV}_t^*(h)$, $\text{RV}_t(h)$, IV_t and $\text{IV}_{t+1:t+m}$, defined by Eqs. (2.4), (2.10), (2.2) and (2.3), respectively. Then for integers $m \geq 1$ and $l \geq 0$, and $h > 0$,*

$$\text{Cov}[\text{IV}_{t+1:t+m}, \text{RV}_{t-l}(h)] = \text{Cov}[\text{IV}_{t+1:t+m}, \text{RV}_{t-l}^*(h)]$$
$$= \text{Cov}[\text{IV}_{t+1:t+m}, \text{IV}_{t-l}], \tag{3.2}$$

$$\text{Var}[\text{RV}_t(h)] = \text{Var}[\text{RV}_t^*(h)]$$
$$+ 2V_u^2\left(\frac{2K_u}{h} - K_u + 1 + 4\frac{E[\sigma_t^2]}{V_u}\right), \tag{3.3}$$

$$\text{Cov}[\text{RV}_{t+1}(h), \text{RV}_t(h)] = \text{Cov}[\text{RV}_{t+1}^*(h), \text{RV}_t^*(h)]$$
$$+ (K_u - 1)V_u^2, \tag{3.4}$$

$$\text{Cov}[\text{RV}_{t+1}(h), \text{RV}_{t-l}(h)] = \text{Cov}[\text{RV}_{t+1}^*(h), \text{RV}_{t-l}^*(h)]. \tag{3.5}$$

The proposition expresses variances and covariances for $\text{RV}_t(h)$ via the counterparts for $\text{RV}_t^*(h)$ along with the noise variance

and kurtosis. The relevant expressions for terms involving $RV_t^*(h)$ appear in ABM (2004). Adapting their notation, we have,

$$\text{Var}[IV_{t+1:t+m}] = 2\sum_{n=1}^{p}\frac{a_n^2}{\lambda_n^2}[\exp(-\lambda_n m) + \lambda_n m - 1], \qquad (3.6)$$

$$\text{Cov}(IV_{t+1:t+m}, IV_{t-l})$$

$$= \sum_{n=1}^{p} a_n^2 \frac{[1 - \exp(-\lambda_n)][1 - \exp(-\lambda_n m)]}{\lambda_n^2} \exp(-\lambda_n l), \qquad (3.7)$$

$$\text{Var}[RV_{t+1}^*(h)] = \text{Var}[IV_{t+1}]$$

$$+ \frac{4}{h}\left(\frac{a_0^2 h^2}{2} + \sum_{n=1}^{p}\frac{a_n^2}{\lambda_n^2}[\exp(-\lambda_n h) - 1 + \lambda_n h]\right), \qquad (3.8)$$

$$\text{Cov}[RV_{t+1}^*(h), RV_{t-l}^*(h)] = \text{Cov}[IV_{t+1}(h), IV_{t-l}(h)]. \qquad (3.9)$$

Combining these expressions with Proposition 3.1, we obtain the requisite variances and covariances for the noise-contaminated realized volatility based forecasts.

We cannot quantify the economic losses due to the adverse impact of microstructure noise on the precision of volatility forecasts, as the appropriate loss function depends on the application. Instead, for compatibility with the extant literature, we focus on the R^2 from the regression of future integrated variance on a constant and the associated forecast variables. This is equivalent to adopting a mean-squared-error (MSE) criterion for the unconditional bias-corrected return variation forecast.[13] The main limitation is that we do not consider (unknown) time-variation in the noise distribution which would reduce the effectiveness of the (unconditional) bias-correction. As such, our analysis provides only a first step towards understanding the impact of noise on volatility forecasting.

The R^2 from the Mincer–Zarnowitz style regression of IV_{t+1} onto a constant and the $(l + 1) \times 1$ vector, $(RV_t(h), RV_{t-1}(h), \dots, RV_{t-l}(h))$, $l \geq 0$, may be succinctly expressed as,

$$R^2(IV_{t+1}, RV_t(h), l) = C(IV_{t+1}, RV_t(h), l)^\top (M(RV_t(h), l))^{-1}$$

$$\times C(IV_{t+1}, RV_t(h), l)/\text{Var}[IV_t] \qquad (3.10)$$

where the typical elements in the $(l + 1) \times 1$ vector $C(IV_{t+1}, RV_t(h), l)$ and the $(l + 1) \times (l + 1)$ matrix $M(RV_t(h), l)$ are given by, respectively,

$$C(IV_{t+1}, RV_t(h), l)_i = \text{Cov}(IV_{t+1}, RV_{t-i+1}(h)) \qquad (3.11)$$

and,

$$M(RV_t(h), l)_{ij} = \text{Cov}(RV_t(h), RV_{t+i-j}(h)). \qquad (3.12)$$

The corresponding R^2 for the longer-horizon integrated volatility forecasts is obtained by replacing IV_{t+1} with $IV_{t+1:t+m}$ in the formulas immediately above.

3.2. Quantifying the impact of market microstructure noise

The impact of microstructure noise is related to the size of the noise variance relative to the daily return variation, conveniently captured by the noise-to-signal ratio, or $\gamma \equiv V_u/E[IV_t]$. Hansen and Lunde (2006) estimate this factor for thirty actively traded stocks during the year 2000 and find values around 0.1%, with most slightly lower. The magnitude of the noise has declined in recent years and is now much lower for many stocks. Consequently, we use 0.1% as the benchmark for a realistic, if not inflated, value for γ.

We also explore the impact of a significantly higher noise-to-signal ratio of 0.5%.

Table 1 reports the population R^2 in Eq. (3.10) from the regression of future integrated volatility on various realized measures across different forecast horizons, data generating processes (models), levels of microstructure noise, and sampling frequencies. As reference we include, in row one, the R^2's for the optimal (infeasible) forecasts based on the exact value of the (latent) volatility state variable(s). The next two rows concern the (infeasible) forecasts based on past daily (latent) integrated volatility and potentially an additional four lags, or a full week, of daily integrated volatilities. The next eleven rows report R^2's for realized volatility based forecasts assuming no noise and sampling frequencies spanning $h = 1/1444$ to $h = 1$, representing 1-min to daily returns in a 24-h market, and we refer to them accordingly.[14] Alternatively, the $h = 1/1440$ frequency reflect 15-s sampling over a 6-h equity trading day.

Row one reveals that Model 1 implies a higher degree of predictability than Model 2. The latter embodies a second, less persistent, factor which reduces the serial correlation in volatility. In rows two and three, we find only a small loss of predictability for forecasts based on the last day's integrated volatility rather than spot volatility, while exploiting a full week of daily integrated volatilities is only marginally helpful.

We now consider realized volatility based forecasts in the ideal case without noise. Rows four to fourteen reveal only a small drop in predictive power for forecasts using measures constructed from 1- and 5-min returns. At lower sampling frequencies, where the return variation measures are less precise, the addition of lagged volatility measures becomes progressively more valuable. The results for twenty daily lags (19 extra) in row fourteen mimic the performance of a well-specified GARCH model, as detailed in ABM (2004).

Finally, turning to the new results for forecasts based on realized volatilities constructed from noisy returns, we first observe only a mild degradation in performance for the realistic case with $\gamma = 0.1\%$. However, for the higher noise level in the bottom part of the table, the performance deteriorates more sharply and using lagged volatility measures is now critical in boosting the predictive power. Second, as anticipated, it is not optimal to estimate the realized volatility with ultra-high frequency returns. At the moderate noise level, the performance is better for 5-min rather than 1-min sampling, and as γ grows further, sampling at the 15- and 30-min levels produces the highest coherence between forecasts and future realizations, because noise increasingly dominates sampling variability as the main source of variation in the realized measures. Further evidence of the benefit from sparse sampling in this context is obtained by comparing the decline in predictability from the $\gamma = 0.1\%$ to the $\gamma = 0.5\%$ scenario for the 5-min ($h = 1/288$) versus 30-min ($h = 1/48$) sampling frequency. One finds a drop in the R^2 from moderate to large noise for $h = 1/288$ at the one-day forecast horizon in Model 1 of about 91% to 72% (92% to 84% if lags are exploited) compared to a drop for $h = 1/48$ of about 82% to 75% (88% to 85% with lags). Thirdly, the importance of exploiting lagged realized volatility measures increases sharply with the noise level. Even for $\gamma = 0.1\%$, the measures based on 30-min returns are quite competitive with those using 5-min sampling once the lagged volatility measures are exploited. In fact, for the higher noise level, the 30-min based measures dominate the 5-min based ones in all scenarios. Hence, within the class of linear realized volatility based forecast procedures, 30-min sampling appears to provide a robust and fairly efficient choice as long as past daily realized measures are also exploited.

---

[13] Patton (2011) provides an interesting discussion concerning the choice of loss function for assessing the performance of alternative volatility forecasts when the latent volatility is observed with noise.

[14] The figures in the first fourteen rows of Table 1 are extracted from Tables 1 through 6 of ABM (2004).

Table 1
R^2 for integrated variance forecasts.

γ (%)	$1/h$	Model	M1			M2		
		Horizon	1	5	20	1	5	20
		R^2(Best)	0.977	0.891	0.645	0.830	0.586	0.338
		$R^2(\text{IV}_t)$	0.955	0.871	0.630	0.689	0.445	0.214
		$R^2(\text{IV}_t, 4)$	0.957	0.874	0.632	0.698	0.446	0.227
0	1440	$R^2(\text{RV}_t^*(h))$	0.950	0.867	0.627	0.679	0.439	0.211
		$R^2(\text{RV}_t^*(h), 4)$	0.951	0.868	0.627	0.685	0.450	0.224
	288	$R^2(\text{RV}_t^*(h))$	0.932	0.851	0.615	0.641	0.414	0.199
		$R^2(\text{RV}_t^*(h), 4)$	0.934	0.852	0.616	0.642	0.429	0.216
	96	$R^2(\text{RV}_t^*(h))$	0.891	0.813	0.588	0.563	0.364	0.175
		$R^2(\text{RV}_t^*(h), 4)$	0.908	0.829	0.599	0.580	0.395	0.202
	48	$R^2(\text{RV}_t^*(h))$	0.836	0.762	0.551	0.476	0.307	0.148
		$R^2(\text{RV}_t^*(h), 4)$	0.883	0.805	0.582	0.519	0.360	0.186
	1	$R^2(\text{RV}_t^*(h))$	0.122	0.111	0.081	0.031	0.020	0.010
		$R^2(\text{RV}_t^*(h), 4)$	0.360	0.329	0.238	0.072	0.054	0.029
		$R^2(\text{RV}_t^*(h), 19)$	0.493	0.450	0.325	0.092	0.074	0.043
0.1	1440	$R^2(\text{RV}_t(h))$	0.896	0.817	0.591	0.547	0.353	0.170
		$R^2(\text{RV}_t(h), 4)$	0.911	0.831	0.601	0.569	0.388	0.199
	288	$R^2(\text{RV}_t(h))$	0.908	0.828	0.599	0.581	0.375	0.181
		$R^2(\text{RV}_t(h), 4)$	0.917	0.837	0.605	0.594	0.402	0.205
	96	$R^2(\text{RV}_t(h))$	0.873	0.797	0.576	0.525	0.339	0.163
		$R^2(\text{RV}_t(h), 4)$	0.899	0.821	0.593	0.553	0.379	0.195
	48	$R^2(\text{RV}_t(h))$	0.821	0.749	0.541	0.450	0.291	0.140
		$R^2(\text{RV}_t(h), 4)$	0.877	0.800	0.578	0.501	0.349	0.182
	1	$R^2(\text{RV}_t(h))$	0.122	0.111	0.081	0.031	0.020	0.010
		$R^2(\text{RV}_t(h), 4)$	0.360	0.328	0.237	0.071	0.053	0.029
		$R^2(\text{RV}_t(h), 19)$	0.492	0.449	0.325	0.091	0.074	0.042
0.5	1440	$R^2(\text{RV}_t(h))$	0.446	0.407	0.294	0.123	0.080	0.038
		$R^2(\text{RV}_t(h), 4)$	0.711	0.649	0.469	0.222	0.164	0.088
	288	$R^2(\text{RV}_t(h))$	0.719	0.656	0.474	0.300	0.194	0.093
		$R^2(\text{RV}_t(h), 4)$	0.837	0.764	0.552	0.395	0.283	0.149
	96	$R^2(\text{RV}_t(h))$	0.772	0.704	0.509	0.365	0.236	0.113
		$R^2(\text{RV}_t(h), 4)$	0.858	0.782	0.566	0.443	0.313	0.165
	48	$R^2(\text{RV}_t(h))$	0.750	0.684	0.495	0.349	0.225	0.108
		$R^2(\text{RV}_t(h), 4)$	0.849	0.775	0.560	0.431	0.306	0.161
	1	$R^2(\text{RV}_t(h))$	0.121	0.110	0.080	0.031	0.020	0.010
		$R^2(\text{RV}_t(h), 4)$	0.357	0.326	0.236	0.070	0.052	0.028
		$R^2(\text{RV}_t(h), 19)$	0.491	0.448	0.324	0.090	0.073	0.042

Note: The table provides the R^2 of the MZ regression related to forecasts of $\text{IV}_{t+1:t+n}$ 1, 5 and 20 days ahead, for models M1 and M2; h indicates the size of the intraday return interval; γ is the noise-to-signal ratio, $\text{Var}[\text{noise}]/E[\text{IV}_t]$. The first row provides the optimal forecast from ABM (2004). The next two rows refer to cases where the explanatory variable is current (row 2) and lagged IV_t (row 3). The last three blocks correspond to cases with current and lagged (4 and 19 lags) daily realized volatility as regressors.

3.3. Optimal sampling frequency

The findings in Section 3.2 suggest exploring the notion of an optimal sampling frequency for $\text{RV}_t(h)$ in terms of maximizing the R^2 for the linear forecasting regressions or, equivalently, minimizing the MSE of the forecasts. This section considers two alternative proposals for choosing h. We focus on one-step-ahead forecasts, but the results are readily extended to longer horizons, as exemplified by our numerical calculations below.

One approach follows Bandi and Russell (2006, 2008), and Aït-Sahalia et al. (2005) who show that the optimal sampling frequency, in terms of minimizing the MSE of $\text{RV}_t(h)$ conditional on the sample path of volatility, may be approximated by,

$$h_t^* \approx (IQ_t/(4V_u^2))^{-1/3}, \tag{3.13}$$

where the integrated quarticity is defined by,

$$IQ_t = \int_{t-1}^{t} \sigma_\tau^4 d\tau. \tag{3.14}$$

However, instead of attempting to estimate the optimal frequency on a period-by-period basis, we follow Bandi and Russell (2006) in replacing the hard-to-estimate one-period integrated quarticity by its unconditional expectation. Hence, we consider,

$$h_1 = (E[IQ_t]/(4V_u^2))^{-1/3}. \tag{3.15}$$

This unconditional counterpart to h_t^* is fairly easy to estimate and implement in practice.

We also consider the frequency which minimizes the variance of $\text{RV}_t(h)$. For motivation, note that the R^2 from the regression of IV_{t+1} on a constant and $\text{RV}_t(h)$ is,

$$R^2 = \frac{\text{Cov}[\text{IV}_{t+1}, \text{RV}_t(h)]^2}{\text{Var}[\text{IV}_t]\text{Var}[\text{RV}_t(h)]} = \frac{\text{Cov}[\text{IV}_{t+1}, \text{IV}_t]^2}{\text{Var}[\text{IV}_t]\text{Var}[\text{RV}_t(h)]} \tag{3.16}$$

where the last equality follows from Proposition 3.1. Hence, maximizing this R^2 is tantamount to minimizing the unconditional variance of $\text{Var}[\text{RV}_t(h)]$, also noted by Ghysels and Sinko (2006). To minimize this variance, we follow Barndorff-Nielsen and Shephard (2002), and Meddahi (2002a), in approximating the unconditional variance of the corresponding non-contaminated realized volatility measure by,

$$\text{Var}[\text{RV}_t^*(h)] \approx \text{Var}[\text{IV}_t] + 2hE[IQ_t]. \tag{3.17}$$

Substituting this expression into the equation for $\text{Var}[\text{RV}_t(h)]$ in Eq. (3.3) yields,

$$\text{Var}[\text{RV}_t^*(h)] \approx \text{Var}[\text{IV}_t] + 2hE[IQ_t]$$
$$+ 2V_u^2\left(\frac{2K_u}{h} - K_u + 1 + 4\frac{E[\sigma_t^2]}{V_u}\right). \tag{3.18}$$

Table 2
R^2 for 'optimally' sampled intraday returns.

γ (%)	$1/h_1$	$1/h_2$	Horizon	1	5	20
Model M1						
0.1	70.8	487	$R^2(RV_t(h_1))$	0.854	0.779	0.563
			$R^2(RV_t(h_1), 4)$	0.891	0.813	0.588
			$R^2(RV_t(h_2))$	0.911	0.832	0.601
			$R^2(RV_t(h_2), 4)$	0.919	0.839	0.607
0.5	24.2	97.3	$R^2(RV_t(h_1))$	0.684	0.624	0.451
			$R^2(RV_t(h_1), 4)$	0.824	0.752	0.544
			$R^2(RV_t(h_2))$	0.772	0.704	0.509
			$R^2(RV_t(h_2), 4)$	0.858	0.782	0.566
Model M2						
0.1	65.3	431	$R^2(RV_t(h_1))$	0.487	0.315	0.151
			$R^2(RV_t(h_1), 4)$	0.527	0.364	0.188
			$R^2(RV_t(h_2))$	0.585	0.378	0.182
			$R^2(RV_t(h_2), 4)$	0.597	0.404	0.206
0.5	22.3	86.2	$R^2(RV_t(h_1))$	0.285	0.184	0.089
			$R^2(RV_t(h_1), 4)$	0.383	0.275	0.146
			$R^2(RV_t(h_2))$	0.365	0.236	0.113
			$R^2(RV_t(h_2), 4)$	0.443	0.314	0.165

Note: The table provides the R^2 of the MZ regression related to forecasts of $IV_{t+1:t+n}$ 1, 5 and 20 days ahead, for models M1 and M2; h indicates the size of the intraday return interval; γ is the noise-to-signal ratio, $Var[noise]/E[IV_t]$. The regressors are current and lagged (4) realized volatility computed from return intervals of length h_1 or h_2. h_1 defined in (3.15) corresponds to the sampling interval suggested by Zhang et al. (2005) and Bandi and Russell (2008). h_2 defined in (3.19) corresponds to the frequency maximizing $R^2(IV_{t+1}, RV_t(h))$.

Minimizing with respect to h produces an alternative candidate sampling frequency,

$$h_2 = (E[IQ_t]/(2V_u^2 K_u))^{-1/2}. \tag{3.19}$$

The relative size of h_1 versus h_2 obviously depends on the magnitude and distribution of the noise term as well as the volatility-of-volatility, or $E[IQ_t]$. Importantly, however, both h_1 and h_2 may be estimated in a model-free fashion by using the higher order sample moments of $RV_t(h)$ based on very finely sampled returns, or small h values, to assess V_u and K_u, along with the use of lower frequency returns to estimate $E[IQ_t]$; see Bandi and Russell (2006, 2008) for further discussion and empirical analysis along these lines.

Table 2 reports approximate optimal sampling frequencies, as represented by h_1 and h_2, for the scenarios in Table 1, along with the resulting population R^2's. Since h_2 directly optimizes an approximation of this quantity, we would expect the associated forecasts to outperform those based on h_1. Nonetheless, the size of the discrepancy is noteworthy. In some cases, the R^2 increases by over 25% and there are always a few percent to be gained by adhering to h_2 rather than h_1. The reason is that h_2 invariably prescribes more frequent sampling than h_1. This finding reflects the pronounced right skew in the distribution of integrated quarticity. Large IQ_t values are associated with high optimal sampling frequencies to offset the increase in discretization error. Hence, averaging the optimal frequency across days, as in the derivation of h_1, ignores the disproportional losses suffered on the most volatile days. In contrast, h_2 minimizes the average squared error and thus adjusts the sampling frequency to accommodate the more extreme days.

Of course, if the cost of a fixed sampling frequency is high, one may seek to vary the sampling frequency based on an initial estimate of the integrated quarticity. However, a comparison of the forecast performance associated with h_2 and moderate noise in Table 2 with that stemming from forecasts derived from realized volatility in the absence of noise in rows 4–7 in Table 1 shows that the loss is quite small. Hence, for these models, it seems more important to pin down a sensible sampling frequency than to vary the intraday return interval from day to day in response to the varying

precision of the volatility measure. This is obviously a comforting finding from a practical perspective.[15]

3.4. Optimally combining intra day returns

The basic realized volatility estimator utilizes a flat weighting scheme in combining the information in intraday returns. This is primarily motivated by the consistency property of the measure for the underlying return variation. Once noise is present, the basic measures become inconsistent even if the sparse estimators only suffer from minor finite sample biases. Moreover, inconsistent measures can provide a sensible basis for predicting the future return variation via forecast regressions which adjust for any systematic (unconditional) bias through the inclusion of a constant term. The main issue for forecast regressors is not their bias but their ability to capture variations in the current realized volatility which typically translates into improved predictive performance. This suggests that we may want to loosen the link between the regressors and realized volatility measures. A natural step is to have the daily return variation proxy be a more flexible function of the intraday squared returns. To this end, we next contrast the predictive ability of optimally combined, or weighted, intraday squared returns to the usual realized volatility measure. The former may, for an optimal choice of the $\alpha(h)$ and $\beta_i(h)$ coefficients, be represented by the regression

$$IV_{t+1} = \alpha(h) + \sum_{i=1}^{1/h} \beta_i(h)(r_{t-1+ih}^{(h)})^2 + \eta_{t+1}(h). \tag{3.20}$$

This regression is difficult to implement in practice due to the large number of parameters, $1+1/h$, but we can readily compute its population counterpart within the ESV setting using Proposition 2.1. The corresponding numerical results are presented in Table 3.

Comparing the results to those in the previous tables, the minor gains obtained by optimal intraday weighting are striking. Even if the improvements are slightly larger for Model 2 than Model 1, they will inevitably be negated, in practice, by the need to estimate the weighting scheme a priori. Of course, the flat weighting of the RV estimator strikes a sensible balance between efficiency and parsimony. However, the above representations only allow for linear weighting of the intraday squared returns. Many modern noise robust estimators involve nonlinear functions of the intraday returns. We explore the forecast potential of some of these estimators below.[16]

4. Robust realized volatility based forecasts

This section investigates to what extent reduced form forecast models based on noise-robust realized variation estimators improve on forecasts constructed from traditional realized volatility measures. In particular, we consider the average and two-scale estimators of Aït-Sahalia et al. (2005), the first-order autocovariance adjusted estimator of Zhou (1996), and the realized kernels of Barndorff-Nielsen et al. (2008a).

---

[15] As indicated previously, one major caveat is that time-variation in the noise distribution, which we do not consider, will render bias-correction less effective. This is most critical for procedures requiring frequent sampling which tend to lower sampling variation but increase bias. Hence, this effect may work to offset some of the advantages of h_2 relative to h_1. Future work should further explore this issue.

[16] The MIDAS scheme of Ghysels et al. (2006) also produces regression-based volatility forecasts using nonlinear functions of lagged intraday absolute returns, but this approach generally does not fall within to the analytical ESV framework.

Table 3
R^2 for optimally combined intraday squared returns.

γ (%)	1/h	Model	M1			M2		
		Horizon	1	5	20	1	5	20
0.1	1440	$R^2(\mathrm{RV}_t(h))$	0.896	0.817	0.591	0.547	0.353	0.170
		R^2(optimal)	0.897	0.819	0.592	0.562	0.359	0.171
	288	$R^2(\mathrm{RV}_t(h))$	0.908	0.828	0.599	0.581	0.376	0.180
		R^2(optimal)	0.910	0.830	0.600	0.600	0.382	0.182
	96	$R^2(\mathrm{RV}_t(h))$	0.873	0.797	0.576	0.525	0.339	0.163
		R^2(optimal)	0.874	0.798	0.577	0.537	0.343	0.164
	48	$R^2(\mathrm{RV}_t(h))$	0.821	0.749	0.541	0.450	0.290	0.140
		R^2(optimal)	0.821	0.749	0.542	0.457	0.293	0.140
0.5	1440	$R^2(\mathrm{RV}_t(h))$	0.446	0.407	0.294	0.123	0.080	0.038
		R^2(optimal)	0.446	0.407	0.294	0.124	0.080	0.038
	288	$R^2(\mathrm{RV}_t(h))$	0.719	0.656	0.474	0.300	0.194	0.093
		R^2(optimal)	0.719	0.656	0.475	0.303	0.195	0.093
	96	$R^2(\mathrm{RV}_t(h))$	0.772	0.704	0.509	0.365	0.235	0.113
		R^2(optimal)	0.772	0.705	0.510	0.369	0.237	0.114
	48	$R^2(\mathrm{RV}_t(h))$	0.750	0.684	0.495	0.349	0.225	0.108
		R^2(optimal)	0.750	0.684	0.495	0.353	0.227	0.109

Note: The table provides the R^2 of the MZ regression related to forecasts of $\mathrm{IV}_{t+1:t+n}$ 1, 5 and 20 days ahead, for models M1 and M2; h indicates the length of the intraday return interval; γ is the noise-to-signal ratio, $\mathrm{Var[noise]}/E[\mathrm{IV}_t]$. $R^2(\mathrm{RV}_t(h))$ is for the case where the regressor is $\mathrm{RV}_t(h)$. R^2(optimal) corresponds to the optimal linear forecast combining $r_{t-1+ih}^{(h)2}$, $i = 1, \ldots, 1/h$.

4.1. Quadratic form representation

We first develop a unified quadratic form representation for the alternative estimators.[17] Let $\underline{h}$ denote the shortest practical intraday interval such that $1/\underline{h}$ is an integer. As before, we let $1/h$ denote the actual number of equally spaced returns used to construct a (sparsely sampled) realized volatility estimator. It is convenient to express each such measure as a quadratic function of the $1/\underline{h} \times 1$ vector of the highest frequency returns. That is,

$$RM_t(h) = \sum_{1 \leq i,j \leq 1/\underline{h}} q_{ij} r_{t-1+i\underline{h}}^{(h)} r_{t-1+j\underline{h}}^{(h)} = R_t(\underline{h})^\top Q R_t(\underline{h}) \qquad (4.1)$$

where the $(1/\underline{h} \times 1)$ vector $R_t(\underline{h})$ is defined by,

$$R_t(\underline{h}) = (r_{t-1+\underline{h}}^{(h)}, r_{t-1+2\underline{h}}^{(h)}, \ldots, r_t^{(h)})^\top. \qquad (4.2)$$

In order to study the interaction between these alternative volatility measures and their relation to the underlying integrated variance, we need analytical expressions for the corresponding first and second moments. The next proposition delivers these quantities.

Proposition 4.1. *Let the noise-contaminated returns be given by Eqs. (2.1) and (2.8), let $\overline{RM}_t(h)$ and $RM_t(h)$ denote two realized volatility measures defined via Eq. (4.1) with corresponding quadratic form weights $\bar{q}_{ij}$ and q_{ij}, and let the integrated volatilities, IV_t and $\mathrm{IV}_{t+1:t+m}$, be defined by Eqs. (2.2)–(2.3). Then,*

$$E[RM_t(h)] = \sum_{1 \leq i,j \leq 1/\underline{h}} q_{ij} E[r_{t-1+i\underline{h}}^{(h)} r_{t-1+j\underline{h}}^{(h)}], \qquad (4.3)$$

$$E[RM_t(h)^2] = \sum_{1 \leq i,j,k,l \leq 1/\underline{h}} q_{ij} q_{kl} E[r_{t-1+i\underline{h}}^{(h)} r_{t-1+j\underline{h}}^{(h)} r_{t-1+k\underline{h}}^{(h)} r_{t-1+l\underline{h}}^{(h)}], \qquad (4.4)$$

$$E[RM_t(h)\overline{RM}_t(h)]$$
$$= \sum_{1 \leq i,j,k,l \leq 1/\underline{h}} q_{ij} \bar{q}_{kl} E[r_{t-1+i\underline{h}}^{(h)} r_{t-1+j\underline{h}}^{(h)} r_{t-1+k\underline{h}}^{(h)} r_{t-1+l\underline{h}}^{(h)}], \qquad (4.5)$$

$$E[\mathrm{IV}_t RM_t(h)] = \sum_{1 \leq i,j \leq 1/\underline{h}} q_{ij} E[\mathrm{IV}_t r_{t-1+i\underline{h}}^{(h)} r_{t-1+j\underline{h}}^{(h)}], \qquad (4.6)$$

$$E[\mathrm{IV}_{t+1:t+m} RM_t(h)] = \sum_{1 \leq i,j \leq 1/\underline{h}} q_{ij} E[\mathrm{IV}_{t+1:t+m} r_{t-1+i\underline{h}}^{(h)} r_{t-1+j\underline{h}}^{(h)}]. \qquad (4.7)$$

Proof. Follows directly from the quadratic form representation. For the ESV model class, closed-form expressions for the right-hand-side of the equations in Proposition 4.1 follow from Proposition 2.1. $\square$

4.2. Robust RV estimators

4.2.1. The "all" RV estimator

The "all" estimator equals the standard realized volatility applied to the maximal sampling frequency. The quadratic form representation is simply,

$$\mathrm{RV}_t^{\mathrm{all}}(\underline{h}) \equiv \mathrm{RV}_t(\underline{h}) = \sum_{i=1}^{1/\underline{h}} r_{t-1+i\underline{h}}^{(h)2} = R_t(\underline{h})^\top Q^{\mathrm{all}}(\underline{h}) R_t(\underline{h}), \qquad (4.8)$$

where

$$q_{ij}^{\mathrm{all}}(\underline{h}) = 1 \quad \text{for } i = j \quad \text{and} \quad q_{ij}^{\mathrm{all}}(\underline{h}) = 0 \quad \text{when } i \neq j. \qquad (4.9)$$

This measure is not noise-robust so it is a poor estimator of IV_t for small $\underline{h}$. However, it plays an important role in defining some of the estimators below.

4.2.2. The sparse RV estimator

The sparse estimator equals the usual $\mathrm{RV}_t(h)$ measure, except that h is a multiple of $\underline{h}$; i.e., $h = \underline{h} n_h$ for n_h a positive integer. The quadratic representation takes the form,

$$\mathrm{RV}_t^{\mathrm{sparse}}(h) = \sum_{i=1}^{1/h} r_{t-1+ih}^{(h)2} = R_t(\underline{h})^\top Q^{\mathrm{sparse}}(h) R_t(\underline{h}), \qquad (4.10)$$

where

$$q_{ij}^{\mathrm{sparse}}(h) = 1 \quad \text{for } i = j, \text{ or for } i \neq j, (s-1)n_h$$
$$+ 1 \leq i,j \leq sn_h, s = 1, \ldots, 1/h,$$
$$= 0 \quad \text{otherwise.} \qquad (4.11)$$

For a larger h, this estimator is more noise robust but will be subject to increased sampling variability. It also serves as a building block for more desirable estimators.

---

[17] Sun (2006) introduced the same quadratic representation independently from this paper.

4.2.3. The average RV estimator

Aït-Sahalia et al. (2005) define the average (or "subsampled") RV estimator as the mean of several sparse estimators. In particular, define the n_h distinct sparse estimators initiated respectively at $0, \underline{h}, 2\underline{h}, \ldots, (n_h - 1)\underline{h}$, through the equation,

$$RV_t^{\text{sparse}}(h, k) = \sum_{i=1}^{N_k} (r_{t-1+k\underline{h}+ih}^{(h)})^2, \quad k = 0, \ldots, n_h - 1, \quad (4.12)$$

where, as before, $h = \underline{h}n_h$, and

$$N_k = \frac{1}{h} \quad \text{if } k = 0, \qquad N_k = \frac{1}{h} - 1 \quad \text{if } k = 1, \ldots, n_h - 1. \quad (4.13)$$

In terms of the quadratic form representation we have,

$$RV_t^{\text{sparse}}(h, k) = \sum_{i=1}^{N_k} (r_{t-1+k\underline{h}+ih}^{(h)})^2 = R_t(\underline{h})^\top Q^{\text{sparse}}(h, k)R_t(\underline{h}),$$

where

$$
\begin{aligned}
q_{ij}^{\text{sparse}}(h, k) &= 1 \quad \text{for } k + 1 \le i = j \le N_k + k, \\
&= 1 \quad \text{for } i \ne j, (s - 1)n_h \\
&\quad + 1 + k \le i, j \le sn_h + k, s = 1, \ldots, N_k/n_h, \\
&= 0 \quad \text{otherwise.} \quad (4.14)
\end{aligned}
$$

The average estimator is now simply defined by the mean of these sparse estimators,

$$
\begin{aligned}
RV_t^{\text{average}}(h) &= \frac{1}{n_h} \sum_{k=0}^{n_h - 1} RV_t^{\text{sparse}}(h, k) \\
&= R_t(\underline{h})^\top Q^{\text{sparse}}(h, k)R_t(\underline{h}), \quad (4.15)
\end{aligned}
$$

where

$$q_{ij}^{\text{average}}(h) = \frac{1}{n_h} \sum_{k=0}^{n_h - 1} q_{ij}^{\text{sparse}}(h, k). \quad (4.16)$$

Whereas the sparse estimator only uses a small subset of the available data, the average estimator exploits more of the data by extending the estimator to each subgrid partition while retaining the associated robustness to noise for appropriately large values of h.

4.2.4. The (adjusted) two-scale RV estimator

The two-scale estimator of Zhang et al. (2005) is obtained by combining the $RV_t^{\text{average}}(h)$ and $RV_t^{\text{all}}(\underline{h})$ estimators. Specifically, let

$$
\begin{aligned}
\bar{n} &\equiv \frac{1}{n_h} \sum_{k=0}^{n_h - 1} N_k = \frac{1}{n_h} \left(\frac{1}{h} + (n_h - 1)\left(\frac{1}{h} - 1 \right) \right) \\
&= \frac{1}{h} - 1 + \frac{h}{\underline{h}}. \quad (4.17)
\end{aligned}
$$

The (finite-sample adjusted) two-scale estimator may then be expressed as,

$$
\begin{aligned}
RV_t^{TS}(h) &= \left(1 - \bar{n}\underline{h} \right)^{-1} \left(RV_t^{\text{average}}(h) - \bar{n}\underline{h}RV_t^{\text{all}} \right) \\
&= R_t(\underline{h})^\top Q^{TS}(h)R_t(\underline{h}), \quad (4.18)
\end{aligned}
$$

where

$$q_{ij}^{TS}(h) = \left(1 - \bar{n}\underline{h} \right)^{-1} (q_{ij}^{\text{average}}(h) - \bar{n}\underline{h}q_{ij}^{\text{all}}(\underline{h})). \quad (4.19)$$

The initial scaling factor provides a simple finite-sample adjustment reflecting the number of terms entering into each of the two summands defining the two-scale estimator.

Unlike the previously defined estimators, the two-scale measure is consistent for IV_t as $h \to 0$ under the noise assumptions in Section 2. Zhang (2006) analyzes related, but more elaborate, multi-scale estimators. We do not consider these extensions here.

4.2.5. Zhou's RV estimator

The estimator originally proposed by Zhou (1996) essentially involves a correction for first-order serial correlation in the high-frequency returns, leading to an unbiased but still inconsistent estimator of the integrated variance. Specifically,

$$
\begin{aligned}
RV_t^{\text{Zhou}}(h) &= \sum_{i=1}^{1/h} r_{t-1+ih}^{(h)2} + \sum_{i=2}^{1/h} r_{t-1+ih}^{(h)} r_{t-1+(i-1)h}^{(h)} \\
&\quad + \sum_{i=1}^{1/h-1} r_{t-1+ih}^{(h)} r_{t-1+(i+1)h}^{(h)} \\
&= R_t(\underline{h})^\top Q^{\text{Zhou}}(h)R_t(\underline{h}), \quad (4.20)
\end{aligned}
$$

where

$$
\begin{aligned}
q_{ij}^{\text{Zhou}}(h) &= 1 \quad \text{for } i = j, \text{ or for } i \ne j, (s - 1)n_h \\
&\quad + 1 \le i, j \le sn_h, s = 1, \ldots, 1/h, \\
&= 1 \quad \text{for } (s - 1)n_h \\
&\quad + 1 \le i, j^* \le sn_h, j^* = j \pm n_h, s = 1, \ldots, 1/h, \\
&= 0 \quad \text{otherwise.} \quad (4.21)
\end{aligned}
$$

Upon defining the estimator at the highest frequency $\underline{h}$, the expressions simplify, as $q_{ij}^{\text{Zhou}}(\underline{h}) = 1$ if $|i - j| \le 1$, and $q_{ij}^{\text{Zhou}}(\underline{h}) = 0$ otherwise. This is the version of the estimator used in deriving the numerical results below. This estimator has also previously been analyzed by Zumbach et al. (2002).

4.2.6. Realized kernels

The Zhou estimator is a special case of the realized kernels developed by Barndorff-Nielsen et al. (2008a). Letting $K(\cdot)$ and L denote the kernel and bandwidth, the realized kernel RV estimator is given by,

$$RV_t^{\text{Kernel}}(K(\cdot), L) = RV_t(\underline{h}) + \sum_{l=1}^{L} K\left(\frac{l-1}{L} \right) CRV_t(l, \underline{h}), \quad (4.22)$$

where

$$CRV_t(l, \underline{h}) = \sum_{i=1+l}^{1/h} r_{t-1+i\underline{h}}^{(h)} r_{t-1+(i-l)\underline{h}}^{(h)} + \sum_{i=1}^{1/h-l} r_{t-1+i\underline{h}}^{(h)} r_{t-1+(i+l)\underline{h}}^{(h)}. \quad (4.23)$$

This estimator is readily expressed in quadratic form as,

$$RV_t^{\text{Kernel}}(K(\cdot), L) = R_t(\underline{h})^\top Q^{\text{Kernel}}(K(\cdot), L)R_t(\underline{h}), \quad (4.24)$$

where

$$
\begin{aligned}
q_{ij}^{\text{Kernel}}(K(\cdot), L) &= 1 \quad \text{for } i = j \\
&= K\left(\frac{l-1}{L} \right) \quad \text{for } |i - j| = l, l \le L, \\
&= 0 \quad \text{otherwise.} \quad (4.25)
\end{aligned}
$$

In the specific calculations below, we use the modified Tukey–Hanning kernel of order two advocated by Barndorff-Nielsen et al. (2008a),[18]

$$K(x) = \sin^2(\pi(1 - x)^2/2). \quad (4.26)$$

We do not use the bandwidth selection procedure of Barndorff-Nielsen et al. (2008a) in our benchmark kernel RV estimator but fix $L = h/\underline{h} - 1$. However, our framework allows for direct comparison of the estimators across bandwidth choices and we explore the performance for a range of reasonable values in Section 4.4.

---

[18] The estimator we implement differs slightly from theirs as, in contrast to Eq. (4.23), they add returns outside the $[t - 1, t]$ time interval to avoid certain end-effects. Since our analysis focuses on forecasting, we want to avoid including any returns beyond time t in the realized volatility measure for the $[t - 1, t]$ interval. This renders the estimator inconsistent although we, a priori, expect the quantitative impact to be minor.

Table 4
Mean, variance and MSE of RV measures.

Model	M1			M2		
IV_t	Mean	Variance	MSE	Mean	Variance	MSE
	0.636	0.168	0.168	0.504	0.0263	0.0263
$\gamma = 0.1\%$						
RV_t^{all}	2.47	0.179	3.53	1.96	0.033	2.14
RV_t^{sparse}	1.002	0.177	0.311	0.795	0.031	0.116
$RV_t^{average}$	1.000	0.171	0.303	0.793	0.028	0.111
RV_t^{TS}	0.634	0.172	0.172	0.503	0.027	0.027
RV_t^{Zhou}	0.637	0.178	0.178	0.505	0.032	0.032
RV_t^{Kernel}	0.637	0.173	0.173	0.506	0.029	0.029
$\gamma = 0.5\%$						
RV_t^{all}	9.79	0.360	84.2	7.77	0.147	52.9
RV_t^{sparse}	2.47	0.223	3.58	1.96	0.060	2.17
$RV_t^{average}$	2.46	0.180	3.51	1.95	0.034	2.13
RV_t^{TS}	0.634	0.182	0.182	0.503	0.035	0.035
RV_t^{Zhou}	0.642	0.303	0.303	0.509	0.111	0.111
RV_t^{Kernel}	0.642	0.194	0.194	0.509	0.042	0.042

Note: The table provides summary statistics for the integrated variance and several realized measures, for models M1 and M2. γ is the noise-to-signal ratio, $Var[noise]/E[IV_t]$. $RV_t^{all} = RV_t(1/1444)$; $RV_t^{sparse} = RV_t(1/288)$; $RV_t^{average}$ averages the five $RV_t^{sparse}(1/288, k)$, $1 \leq k \leq 5$, measures; RV^{TS} is the adjusted two-scale measure combining the RV_t^{all} and $RV_t^{average}$ defined above; $RV_t^{Zhou} = RV_t^{Zhou}(1/1444)$ from (4.21); $RV_t^{Kernel} = RV_t^{Kernel}$(Tukey–Hanning, 4) given by (4.25).

4.3. Distribution of robust RV measures

The analytical solution for relevant cross-moments within the ESV class facilitates comparison of the properties of the estimators, even in the presence of noise.

Table 4 reports the mean, variance and mean-squared-error for alternative measures of integrated variance. In principle, the "all" estimator employs the highest possible frequency. We fix $\underline{h} = 1/1440$, or 1-min (15-s) sampling in a 24-h (6-h) market, as the shortest practical return interval. As predicted, the "all" estimator is severely inflated by microstructure noise. Under moderate noise, the estimator is, on average, almost four times as large as the underlying integrated variance while this factor rises to ten at the larger noise level, so the measure is useless as a direct estimator for the integrated variance. Moving to the sparse estimator based on $h = 1/288$, or 5-min sampling in a 24-h market, the upward bias remains large although it has dropped sharply relative to the "all" estimator. Reducing the sampling frequency further produces an even less biased estimator but we retain this relatively high frequency to explore more cleanly the implications of the noise-induced bias for the predictive ability of these measures compared to the more robust ones discussed below. The last estimator constructed directly from the standard realized volatility measure is the average estimator appearing in the third row. The averaging reduces the sampling variability, and in turn provides an improvement in the MSE compared to the sparse estimator.

The noise-robust estimators are all virtually unbiased for both models and noise levels, even if we have not optimized the sampling frequency or bandwidth but keep them fixed across all model designs.[19] For Model 1 and moderate noise, the estimators have close to identical sampling variability, but for all other scenarios the average, two-scale and kernel measures display the lowest variability. In particular, the Zhou estimator is not competitive in this regard even though it is designed explicitly for this type of noise structure.

Table 5
Correlations of RV measures.

Model M2	IV_t	RV_t^{all}	RV_t^{sparse}	$RV_t^{average}$	RV_t^{TS}	RV_t^{Zhou}	RV_t^{Kernel}
$\gamma = 0.1\%$							
IV_t	1.00	0.891	0.918	0.965	0.954	0.900	0.953
RV_t^{all}	–	1.00	0.861	0.905	0.851	0.769	0.843
RV_t^{sparse}	–	–	1.00	0.951	0.945	0.872	0.934
$RV_t^{average}$	–	–	–	1.00	0.994	0.916	0.981
RV_t^{TS}	–	–	–	–	1.00	0.926	0.987
RV_t^{Zhou}	–	–	–	–	–	1.00	0.940
RV_t^{Kernel}	–	–	–	–	–	–	1.00
$\gamma = 0.5\%$							
IV_t	1.00	0.423	0.660	0.878	0.863	0.487	0.792
RV_t^{all}	–	1.00	0.460	0.613	0.243	−0.078	0.153
RV_t^{sparse}	–	–	1.00	0.751	0.688	0.359	0.590
$RV_t^{average}$	–	–	–	1.00	0.915	0.477	0.787
RV_t^{TS}	–	–	–	–	1.00	0.626	0.889
RV_t^{Zhou}	–	–	–	–	–	1.00	0.653
RV_t^{Kernel}	–	–	–	–	–	–	1.00

Note: The table provides cross-correlation of the integrated variance and selected realized measures for model M2. γ is the noise-to-signal ratio, $Var[noise]/E[IV_t]$. $RV_t^{all} = RV_t(1/1444)$; $RV_t^{sparse} = RV_t(1/288)$; $RV_t^{average}$ averages the five $RV_t^{sparse}(1/288, k)$, $1 \leq k \leq 5$, measures; RV^{TS} is the adjusted two-scale measure combining the RV_t^{all} and $RV_t^{average}$ defined above; $RV_t^{Zhou} = RV_t^{Zhou}(1/1444)$ from (4.21); $RV_t^{Kernel} = RV_t^{Kernel}$(Tukey–Hanning, 4) given by (4.25).

Table 5 provides the correlations among the alternative estimators as well as the actual integrated variance. This provides a first impression of the potential forecast performance, as high correlation with the current volatility level, everything else equal, should translate into a good prediction. Overall, the measures separate into two distinct groups. The "all", sparse and Zhou estimators fail to match the performance of the remainder in terms of coherence with the ideal integrated variance measure. The average estimator performs well in spite of its sizeable bias, while the nearly unbiased Zhou estimator is handicapped by its larger sampling variability and fails dramatically in the more noisy scenarios. Finally, we stress that the TS and kernel estimators are loosely calibrated to Model 2 with moderate noise, so the entries for these estimators across the other scenarios are less telling.

We have also explored the population autocorrelation of the alternative estimators. Intuitively, the less noisy estimators manage to correlate better with the integrated variance and they may thus be expected to inherit the strong serial dependence present in the daily integrated variance series. This is exactly what we find: the ranking in terms of high correlation with the integrated variance measure in Table 5 is preserved when ranking the estimators in terms of serial dependence. These findings are tabulated in ABM (2006).

4.4. True forecast performance of robust RV measures

We now compare the potential performance of linear forecasts constructed from the alternative return variation measures. In a direct extension of the findings for the regular RV measures in Table 1, we compute the true population R^2's by combining the results from Propositions 2.1 and 4.1. Given the wide array of alternatives, we focus on only one version of each estimator. Hence, the estimators are not calibrated optimally for each scenario but are, at best, designed to perform well for a couple of the relevant cases. Nonetheless, the results are sufficiently impressive that further improvements are unlike to alter the qualitative conclusions.

Table 6 provides the results for daily, weekly and monthly forecast horizons. As expected, the measures most highly correlated with the true return variation also provide the best basis for forecasts. Hence, forecasts generated by the average estimator

---

[19] As illustrated in Section 4.4 below, the frequency and bandwidth are close to optimal for the two-scale and kernel estimator, respectively, in the empirically relevant scenario of Model 2 and moderate noise.

Table 6
R^2 for integrated variance forecasts.

γ (%)	Model	M1			M2		
	Horizon	1	5	20	1	5	20
0.1	$R^2(\mathrm{RV}_t^{\mathrm{all}})$	0.896	0.817	0.591	0.547	0.353	0.170
	$R^2(\mathrm{RV}_t^{\mathrm{sparse}})$	0.908	0.829	0.599	0.581	0.375	0.181
	$R^2(\mathrm{RV}_t^{\mathrm{average}})$	0.934	0.852	0.616	0.642	0.415	0.199
	$R^2(\mathrm{RV}_t^{\mathrm{TS}})$	0.927	0.846	0.612	0.628	0.405	0.195
	$R^2(\mathrm{RV}_t^{\mathrm{Zhou}})$	0.900	0.821	0.593	0.559	0.361	0.174
	$R^2(\mathrm{RV}_t^{\mathrm{kernel}})$	0.928	0.846	0.612	0.626	0.404	0.194
0.5	$R^2(\mathrm{RV}_t^{\mathrm{all}})$	0.446	0.407	0.294	0.123	0.080	0.038
	$R^2(\mathrm{RV}_t^{\mathrm{sparse}})$	0.719	0.656	0.474	0.300	0.194	0.093
	$R^2(\mathrm{RV}_t^{\mathrm{average}})$	0.886	0.809	0.585	0.532	0.343	0.165
	$R^2(\mathrm{RV}_t^{\mathrm{TS}})$	0.876	0.799	0.578	0.513	0.331	0.159
	$R^2(\mathrm{RV}_t^{\mathrm{Zhou}})$	0.529	0.483	0.349	0.163	0.106	0.051
	$R^2(\mathrm{RV}_t^{\mathrm{kernel}})$	0.829	0.756	0.547	0.432	0.279	0.134

Note: The table provides the R^2 of the MZ regression related to forecasts of $\mathrm{IV}_{t+1:t+n}$ 1, 5 and 20 days ahead, for models M1 and M2; γ is the noise-to-signal ratio, $\mathrm{Var[noise]}/E[\mathrm{IV}_t]$. The explanatory variable is one realized measure. $\mathrm{RV}_t^{\mathrm{all}} = \mathrm{RV}_t(1/1444)$; $\mathrm{RV}_t^{\mathrm{sparse}} = \mathrm{RV}_t(1/288)$; $\mathrm{RV}_t^{\mathrm{average}}$ averages the five measures $\mathrm{RV}_t^{\mathrm{sparse}}(1/288, k)$, $1 \leq k \leq 5$; $\mathrm{RV}^{\mathrm{TS}}$ is the adjusted two-scale measure combining $\mathrm{RV}_t^{\mathrm{all}}$ and $\mathrm{RV}_t^{\mathrm{average}}$ defined above; $\mathrm{RV}_t^{\mathrm{Zhou}} = \mathrm{RV}_t^{\mathrm{Zhou}}(1/1444)$ defined in (4.21); $\mathrm{RV}_t^{\mathrm{Kernel}} = \mathrm{RV}_t^{\mathrm{Kernel}}$(Tukey–Hanning, 4) given by (4.25).

are uniformly best, although barely distinguishable from those based on the two-scale or kernel estimators when these are well calibrated. Moreover, the fall-off for the remaining forecasts is not dramatic under the realistic moderate noise setting. In fact, compared with the feasible estimators under ideal noise-free conditions, provided in Table 1, the performance of the entire range of forecasts is quite impressive. However, the picture changes for the noisier scenario where sizeable gains are attained by adhering to forecasts based on estimators which succeed in dampening the impact of the noise on the sampling variability.[20] Overall, the evidence suggests that the comparatively simple average (or "sub-sampled") RV estimator is an excellent starting point for practical volatility forecasting.[21]

We have informally stressed the importance of low sampling variability for forecast performance across the alternative realized variation measures. This intuition may be formalized through an extension of the result underlying Eq. (3.16).

Proposition 4.2. *Let the discrete-time noise contaminated returns be determined by an ESV model and the relationship in Eq.* (2.8)*. Let $RM_t(h)$ denote a realized volatility measure as defined in Eq.* (4.1) *with corresponding quadratic form weights q_{ij} such that*

$$\forall i, \; 1 \leq i \leq 1/\underline{h}, \quad q_{ii} = 1. \tag{4.27}$$

Then,

$$\mathrm{Cov}[\mathrm{IV}_{t+1}, RM_t(h)] = \mathrm{Cov}[\mathrm{IV}_{t+1}, \mathrm{IV}_t]. \tag{4.28}$$

Consequently, maximizing the R^2 from the regression of IV_{t+1} on a constant and a $RM_t(h)$ measure of the form (4.1) under the restriction (4.27) is tantamount to minimizing the variance of the

---

[20] In this scenario, the sparse estimator performs better at a lower frequency, say $h = 1/96$, and if lagged measures are used to form the forecasts. However, a lower sampling frequency increases sampling variability so the approach cannot match the more elaborate noise-robust procedures.

[21] The superior performance of the average estimator may be an artifact of the *i.i.d.* noise assumption although we expect it also to perform well for dependent noise. It is possible to accommodate more complex noise structures within a tractable ESV framework, as discussed in ABM (2006), but the lack of consensus regarding the dependence structure in the noise process is an obstacle for broadly exploring the issue.

Table 7
R^2 for one-step ahead forecast of integrated variance when the bandwidth varies.

γ (%)	Model	M1			M2		
	Bandwidth	Average	TS	Kernel	Average	TS	Kernel
0.1	1	0.929	0.899	0.900	0.442	0.143	0.559
	2	0.935	0.924	0.912	0.527	0.295	0.583
	3	0.935	0.928	0.925	0.587	0.531	0.615
	4	0.934	0.927	0.928	0.642	0.628	0.626
	6	0.930	0.924	0.928	0.635	0.623	0.629
	9	0.923	0.917	0.925	0.621	0.610	0.623
	11	0.918	0.912	0.922	0.611	0.601	0.618
	14	0.910	0.904	0.917	0.597	0.586	0.609
0.5	1	0.728	0.530	0.529	0.138	0.012	0.163
	2	0.823	0.771	0.628	0.295	0.069	0.211
	3	0.867	0.846	0.769	0.428	0.321	0.338
	4	0.886	0.876	0.829	0.532	0.513	0.432
	6	0.903	0.897	0.878	0.569	0.559	0.517
	9	0.908	0.903	0.900	0.583	0.575	0.565
	11	0.906	0.901	0.904	0.583	0.575	0.576
	14	0.902	0.897	0.905	0.577	0.568	0.580

Note: The table provides the R^2 of the MZ regression related to forecasts of $\mathrm{IV}_{t+1:t+n}$ 1, 5 and 20 days ahead, for models M1 and M2; γ is the noise-to-signal ratio, $\mathrm{Var[noise]}/E[\mathrm{IV}_t]$. The explanatory variable is either $\mathrm{RV}_t^{\mathrm{average}}$, $\mathrm{RV}_t^{\mathrm{TS}}$ or $\mathrm{RV}_t^{\mathrm{kernel}}$. For a given bandwidth L, $\mathrm{RV}_t^{\mathrm{average}}$ averages the $(L + 1)$ measures $\mathrm{RV}_t^{\mathrm{sparse}}((L + 1)/1444, k)$, $1 \leq k \leq L + 1$; $\mathrm{RV}^{\mathrm{TS}}$ is the adjusted two-scale measure that combines $\mathrm{RV}_t^{\mathrm{all}} = \mathrm{RV}_t(1/1444)$ and $\mathrm{RV}_t^{\mathrm{average}}$ with the same bandwidth; $\mathrm{RV}_t^{\mathrm{Kernel}} = \mathrm{RV}_t^{\mathrm{Kernel}}$(Tukey–Hanning, L) defined in (4.25).

measure $RM_t(h)$. The restriction (4.27) holds for the sparse and Zhou estimators, and for any kernel estimator including the non-flat top kernels introduced by Barndorff-Nielsen et al. (2008b). It is not satisfied for the average and two scale estimators at the edges of the trading day, although it will be close to valid for these as well in most circumstances.

Finally, we explore the importance of calibrating the sampling frequency and bandwidth for the average, TS and kernel estimators. Table 7 reports the population R^2 across model designs for bandwidths spanning 1 to 14. Evidently, higher noise levels and less persistent volatility processes (Model 2) tend to increase the optimal bandwidth. Moreover, there is a distinct pattern to the degree of predictability as the bandwidth rises: performance improves, then levels off and declines. Only for the kernels in the high noise scenario do we not observe a maximum degree of predictability, as this noise level is best accommodated with a very conservative bandwidth. Note also that a bandwidth of four, as in Tables 4–6, is close to optimal for all the estimators in the more realistic scenario of Model 2 and moderate noise.

4.5. Feasible forecasting performance of robust RV measures

The integrated volatility regressand of the Mincer–Zarnowitz regressions in Section 4.4 is latent. In practice it must be replaced by some realized volatility measure, as in,

$$\overline{RM}_{t+1:t+m}(h) = a + bRM_t(h) + \eta_{t+m}, \tag{4.29}$$

where $\overline{RM}_t(h)$ and $RM_t(h)$ denote possibly different realized measures. The associated regression R^2 involves a covariance term which we have not directly considered previously,

$$R^2 = \frac{(\mathrm{Cov}[\overline{RM}_{t+1:t+m}(h), RM_t(h)])^2}{\mathrm{Var}[\overline{RM}_{t+1:t+m}(h)]\mathrm{Var}[RM_t(h)]}. \tag{4.30}$$

The following proposition provides closed form expressions for the requisite covariance term.

Proposition 4.3. *Let the discrete-time noise contaminated returns be determined by an ESV model and the relationship in Eq.* (2.8)*. Let $\overline{RM}_t(h)$ and $RM_t(h)$ denote two realized volatility measures as defined*

Table 8
R^2 for one-step-ahead RV forecasts.

Indp.Var.	RV_t^{all}	RV_t^{sparse}	$RV_t^{average}$	RV_t^{TS}	RV_t^{Zhou}	RV_t^{Kernel}
Model M1						
$\gamma = 0.1\%$						
RV^{all}	0.841	0.852	0.877	0.870	0.844	0.870
$RV^{sp.}$	0.852	0.864	0.889	0.882	0.856	0.882
$RV^{av.}$	0.877	0.889	0.914	0.908	0.880	0.901
RV^{TS}	0.870	0.882	0.908	0.901	0.874	0.901
RV^{Zhou}	0.844	0.856	0.880	0.874	0.848	0.874
$RV^{Ker.}$	0.870	0.882	0.908	0.902	0.874	0.901
$\gamma = 0.5\%$						
RV^{all}	0.208	0.336	0.414	0.409	0.247	0.387
$RV^{sp.}$	0.336	0.542	0.668	0.659	0.399	0.624
$RV^{av.}$	0.414	0.668	0.823	0.813	0.492	0.769
RV^{TS}	0.409	0.659	0.813	0.803	0.486	0.760
RV^{Zhou}	0.247	0.399	0.492	0.486	0.294	0.460
$RV^{Ker.}$	0.387	0.624	0.769	0.760	0.460	0.720
Model M2						
$\gamma = 0.1\%$						
RV^{all}	0.434	0.461	0.510	0.498	0.444	0.497
$RV^{sp.}$	0.461	0.490	0.541	0.529	0.471	0.528
$RV^{av.}$	0.510	0.541	0.598	0.585	0.520	0.588
RV^{TS}	0.498	0.529	0.585	0.572	0.509	0.570
RV^{Zhou}	0.444	0.471	0.520	0.509	0.453	0.507
$RV^{Ker.}$	0.497	0.528	0.583	0.570	0.507	0.568
$\gamma = 0.5\%$						
RV^{all}	0.022	0.054	0.095	0.092	0.029	0.077
$RV^{sp.}$	0.054	0.131	0.231	0.223	0.071	0.188
$RV^{av.}$	0.095	0.231	0.410	0.396	0.126	0.333
RV^{TS}	0.092	0.223	0.396	0.382	0.122	0.321
RV^{Zhou}	0.029	0.071	0.126	0.122	0.039	0.102
$RV^{Ker.}$	0.077	0.188	0.333	0.321	0.102	0.271

Note: The table provides the R^2 of the MZ regression related to forecasts for a realized measure (dependent variable) one-day ahead, for models M1 and M2; γ is the noise-to-signal ratio, $\text{Var}[\text{noise}]/E[\text{IV}_t]$. The explanatory variable is a realized measure. $RV_t^{all} = RV_t(1/1444)$; $RV_t^{sparse} = RV_t(1/288)$; $RV_t^{average}$ averages the five measures $RV_t^{sparse}(1/288, k)$, $1 \leq k \leq 5$; RV^{TS} is the adjusted two-scale measure that combines RV_t^{all} and $RV_t^{average}$ defined above; $RV_t^{Zhou} = RV_t^{Zhou}(1/1444)$ is defined in (4.21); $RV_t^{Kernel} = RV_t^{Kernel}$ (Tukey–Hanning, 4) is given by (4.25).

in Eq. (4.1) with corresponding quadratic form weights q_{ij} and $\bar{q}_{ij}$, respectively. Then,

$$\text{Cov}[\overline{RM}_{t+1}(h), RM_t(h)]$$
$$= \sum_{i=1}^{1/\underline{h}} \sum_{k=1}^{1/\underline{h}} \bar{q}_{ii} q_{kk} \left(\sum_{n=1}^{p} \frac{a_n^2}{\lambda_n^2} \left(1 - \exp(\lambda_n \underline{h})\right)^2 \right.$$
$$\left. \times \exp(-\lambda_n (1 + (i-k-1)\underline{h})) \right)$$
$$+ \bar{q}_{11} q_{\underline{h}^{-1}\underline{h}^{-1}} (K_u - 1) V_u^2, \tag{4.31}$$

and for $m > 1$,

$$\text{Cov}[\overline{RM}_{t+m}(h), RM_t(h)]$$
$$= \sum_{i=1}^{1/\underline{h}} \sum_{k=1}^{1/\underline{h}} \bar{q}_{ii} q_{kk} \left(\sum_{n=1}^{p} \frac{a_n^2}{\lambda_n^2} \left(1 - \exp(\lambda_n \underline{h})\right)^2 \right.$$
$$\left. \times \exp(-\lambda_n (m + (i-k-1)\underline{h})) \right). \tag{4.32}$$

Table 8 provides feasible performance measures that, ideally, may be obtained from the forecast procedures discussed in Section 4.4. We only report figures for the one-step-ahead forecasts

due to the increased number of cross-comparisons.[22] As before, the relative rankings are preserved over the longer horizons. Table 8 conveys the now familiar picture. The average estimator dominates uniformly both as the basis for forecasts and as the proxy for the future realized return variation. Moreover, the prior rankings are preserved everywhere across all the model designs. It is evident that using a precise ex-post estimator for the integrated variance improves the measured degree of predictability and allows the regressions to better convey the true relationship as captured in Table 7. For example, consider Model 2 and $\gamma = 0.5\%$. Using the average estimator as the basis for the forecast and as the ex-post proxy for future return variation realizations, the R^2 is 41% compared to the actual R^2 of about 53% in Table 7. In contrast, exploiting the Zhou estimator in both capacities results in an R^2 below 4%. Although the figures reflect the specific model design, it exemplifies how the issue of observed versus true underlying predictability is crucially important in properly interpreting empirical studies in this area.[23]

5. Conclusion

This paper extends existing analytic methods for the construction and assessment of volatility forecasts for diffusion models to the important case of market microstructure noise. The procedures are valid within the ESV model class, which includes most popular volatility diffusions, and may be adapted to accommodate other empirically relevant features. We apply the techniques to a few representative specifications for which we compare the performance of feasible linear forecasts constructed from alternative realized variation measures in the presence of noise to those based on optimal (infeasible) forecasts. We find it feasible to construct fairly precise forecasts but many aspects of the implementation require careful examination of the underlying market structure and data availability in order to design effective procedures.

Given the vast diversity in potential models, sampling frequencies, levels of microstructure noise, realized variation estimators and forecasting schemes, the costs associated with comprehensive simulation studies are formidable. Instead, the ESV analytical tools developed here enable us to study the relevant issues succinctly across alternative designs within a coherent framework, thus providing a guide for general performance and robustness. As such, we expect the approach to provide additional useful insights in future work concerning the design of alternative return variation measures and their application in the context of volatility forecasting.

Appendix. Technical proofs

Proof of Proposition 2.1. In the absence of any drift, $E[r_{t+ih}^{*(h)}] = 0$ and $\text{Var}[r_{t+ih}^{*(h)}] = a_0 h$ (see, e.g. Meddahi, 2002b). Now given the *i.i.d.* assumption for the noise u_t, (2.15) and (2.16) follows readily from (2.8). Likewise, the non-contaminated returns $r_{t+ih}^{*(h)}$ are uncorrelated (see, e.g. Meddahi, 2002b), while $e_t^{(h)}$ is an MA(1) process. Hence, the observed returns $r_{t+ih}^{(h)}$ will also follow an MA(1) process with

$$\text{Cov}[r_{t+ih}^{*(h)}, r_{t+(i-1)h}^{*(h)}] = \text{Cov}[e_{t+ih}^{(h)}, e_{t+(i-1)h}^{(h)}] = -\text{Var}[u_t] = -V_u,$$

---

[22] Also note that for the feasible regressions analyzed here minimizing the variance of the explanatory robust measure is *not* equivalent to maximizing the R^2 in the Mincer–Zarnowitz regression, as the numerator in the R^2 will also depend on h, and $\text{Cov}[\text{IV}_{t+1}, RM_t(h)] \neq \text{Cov}[\text{IV}_{t+1}, \text{IV}_t]$ for $q_{ii} \neq 1$.

[23] As previously noted, ABM (2005) provide a technique for formally converting the observed degree of predictability into an estimate of the higher true predictability through a fairly simple procedure.

i.e., (2.17). We will now prove (2.18). As a short-hand notation, let r_i, r_i^*, and e_i refer to $r_{t+ih}^{(h)}$, $r_{t+ih}^{*(h)}$, and $e_{t+ih}^{(h)}$. We then have

$$r_i r_j r_k r_l = (r_i^* + e_i)(r_j^* + e_j)(r_k^* + e_k)(r_l^* + e_l)$$
$$= r_i^* r_j^* r_k^* r_l^* + r_i^* r_j^* e_k r_l^* + r_i^* r_j^* r_k^* e_l + r_i^* r_j^* e_k e_l + r_i^* e_j r_k^* r_l^*$$
$$+ r_i^* e_j e_k r_l^* + r_i^* e_j r_k^* e_l + r_i^* e_j e_k e_l + e_i r_j^* r_k^* r_l^* + e_i r_j^* e_k r_l^*$$
$$+ e_i r_j^* r_k^* e_l + e_i r_j^* e_k e_l + e_i e_j r_k^* r_l^*$$
$$+ e_i e_j e_k r_l^* + e_i e_j r_k^* e_l + e_i e_j e_k e_l.$$

The returns are independent from the noise. In addition, the mean of the noise and returns are zero. This implies that quantities like $E[r_i^* r_j^* e_k r_l^*]$ and $E[r_i^* e_j e_k e_l]$ equal zero. Therefore,

$$E[r_i r_j r_k r_l] = E[r_i^* r_j^* r_k^* r_l^*] + E[r_i^* r_j^*]E[e_k e_l] + E[r_i^* r_l^*]E[e_j e_k]$$
$$+ E[r_i^* r_k^*]E[e_j e_l] + E[e_i e_k]E[r_j^* r_l^*] + E[e_i e_l]E[r_j^* r_k^*]$$
$$+ E[e_i e_j]E[r_k^* r_l^*] + E[e_i e_j e_k e_l]. \tag{A.1}$$

We will now compute the elements that appear in (A.1). We start with the first term. Given the path of the volatility, the returns are independent. Therefore (see, e.g. Meddahi, 2002b),

$$E[r_i^* r_j^* r_k^* r_l^*] = E[(r_i^*)^4] \quad \text{if } i = j = k = l,$$
$$= \text{Cov}[(r_{t-1+ih}^{*(h)})^2, (r_{t-1+kh}^{*(h)})^2]$$
$$+ (E[(r_i^*)^2])^2 \quad \text{if } i = j > k = l,$$
$$= 0 \quad \text{otherwise.}$$

Eqs. (3.7) and (3.10) in Meddahi (2002b) now imply

$$E[r_i^* r_j^* r_k^* r_l^*] = 3a_0^2 h^2 + 6\sum_{i=1}^{p} \frac{a_i^2}{\lambda_i^2}[-1 + \lambda_i h + \exp(-\lambda_i h)]$$
$$\text{if } i = j = k = l,$$
$$= \sum_{n=1}^{p} \frac{a_n^2}{\lambda_n^2}[1 - \exp(-\lambda_k h)]^2 \exp(-\lambda_n(i-k-1)h)$$
$$+ a_0^2 h^2 \quad \text{if } i = j > k = l,$$
$$= 0 \quad \text{otherwise.} \tag{A.2}$$

To compute the last term in (A.1) note that

$$e_i e_j e_k e_l = (u_i - u_{i-1})(u_j - u_{j-1})(u_k - u_{k-1})(u_l - u_{l-1})$$
$$= u_i u_j u_k u_l - u_i u_j u_k u_{l-1} - u_i u_j u_{k-1} u_l + u_i u_j u_{k-1} u_{l-1}$$
$$- u_i u_{j-1} u_k u_l + u_i u_{j-1} u_k u_{l-1} + u_i u_{j-1} u_{k-1} u_l$$
$$- u_i u_{j-1} u_{k-1} u_{l-1} - u_{i-1} u_j u_k u_l + u_{i-1} u_j u_k u_{l-1}$$
$$+ u_{i-1} u_j u_{k-1} u_l - u_{i-1} u_j u_{k-1} u_{l-1} + u_{i-1} u_{j-1} u_k u_l$$
$$- u_{i-1} u_{j-1} u_k u_{l-1} - u_{i-1} u_{j-1} u_{k-1} u_l + u_{i-1} u_{j-1} u_{k-1} u_{l-1}.$$

Now, using the *i.i.d.* structure of the u_t process it follows that

$$E[e_i e_j e_k e_l] = 2V_u^2(K_u + 3) \quad \text{if } i = j = k = l,$$
$$= -V_u^2(K_u + 3) \quad \text{if } i = j = k = l + 1$$
$$\text{or } i = j + 1 = k + 1 = l + 1,$$
$$= V_u^2(K_u + 3) \quad \text{if } i = j = k + 1 = l + 1,$$
$$= 4V_u^2 \quad \text{if } i = j > k + 1, k = l,$$
$$= 2V_u^2 \quad \text{if } i = j + 1, j = k = l + 1,$$
$$= -2V_u^2 \quad \text{if } i = j \geq k + 1, k = l + 1$$
$$\text{or } i = j + 1, j \geq k + 1, k = l,$$
$$= V_u^2 \quad \text{if } i = j + 1, j \geq k + 1, k = l + 1,$$
$$= 0 \quad \text{otherwise.} \tag{A.3}$$

Denote the remaining terms that appear in (A.1),

$$A_{ijkl} \equiv E[r_i^* r_j^*]E[e_k e_l] + E[r_i^* r_l^*]E[e_j e_k] + E[r_i^* r_k^*]E[e_j e_l]$$
$$+ E[e_i e_k]E[r_j^* r_l^*] + E[e_i e_l]E[r_j^* r_k^*] + E[e_i e_j]E[r_k^* r_l^*].$$

From above $E[r_i^* r_j^*] = \delta_{i,j}E[(r_i^*)^2] = \delta_{i,j}a_0 h$ and $E[e_i e_j] = \delta_{i,j}E[e_i^2] - \delta_{|i-j|,1}V_u = 2\delta_{i,j}V_u - \delta_{|i-j|,1}V_u$. Hence,

$$A_{ijkl} = 6E[(r_i^*)^2]E[e_i^2] = 12a_0 V_u h \quad \text{if } i = j = k = l,$$
$$= 3E[(r_i^*)^2]E[e_i e_{i-1}] = -3a_0 V_u h \quad \text{if } i = j = k = l + 1$$
$$\text{or } i = j + 1 = k + 1 = l + 1,$$
$$= 2E[(r_i^*)^2]E[e_k^2] = 4a_0 V_u h \quad \text{if } i = j = k + 1 = l + 1,$$
$$= 2E[(r_i^*)^2]E[e_k^2] = 4a_0 V_u h \quad \text{if } i = j > k + 1, k = l,$$
$$= 0 \quad \text{if } i = j + 1, j = k = l + 1,$$
$$= E[(r_i^*)^2]E[e_k e_{k-1}] = -a_0 V_u h \quad \text{if } i = j \geq k + 1, k = l + 1$$
$$\text{or } i = j + 1, j \geq k + 1, k = l,$$
$$= 0 \quad \text{if } i = j + 1, j \geq k + 1, k = l + 1,$$
$$= 0 \quad \text{otherwise.} \tag{A.4}$$

Now combining (A.2)–(A.4) results in (2.18).

The proof of (2.19) proceeds similarly to one for (2.18). The main difference stems from the fact that $t - 1 + m + jh > t - 1 + kh$, so that several terms that appear in (2.18) are now zero. In particular, by using the MA(1) structure of $e_t^{(h)}$, it follows that

$$\text{Cov}[r_{t-1+m+ih}^{(h)} r_{t-1+m+jh}^{(h)}, r_{t-1+kh}^{(h)} r_{t-1+lh}^{(h)}]$$
$$= \text{Cov}[r_{t-1+m+ih}^{*(h)} r_{t-1+m+jh}^{*(h)}, r_{t-1+kh}^{*(h)} r_{t-1+lh}^{*(h)}]$$
$$+ \text{Cov}[e_{t-1+m+ih}^{(h)} e_{t-1+m+jh}^{(h)}, e_{t-1+kh}^{(h)} e_{t-1+lh}^{(h)}]$$
$$= \delta_{i,j}\delta_{k,l}\text{Cov}[(r_{t-1+m+ih}^{*(h)})^2, (r_{t-1+kh}^{*(h)})^2] + \delta_{m,1}\delta_{i,j}\delta_{k,l}\delta_{i,1}\delta_{k,1/h}$$
$$\times \text{Cov}[e_{t+h}^{(h)} e_{t+h}^{(h)}, e_t^{(h)} e_t^{(h)}]$$
$$= \delta_{i,j}\delta_{k,l}\text{Cov}\left[\int_{t-1+m+(i-1)h}^{t-1+m+ih} \sigma_u^2 du, \int_{t-1+(k-1)h}^{t-1+kh} \sigma_u^2 du\right]$$
$$+ \delta_{m,1}\delta_{i,j}\delta_{k,l}\delta_{i,1}\delta_{k,1/h}\text{Cov}[u_t^2, u_t^2]$$
$$= \delta_{i,j}\delta_{k,l}\sum_{n=1}^{p} \frac{a_n^2}{\lambda_n^2}[1 - \exp(-\lambda_n h)]^2 \exp(-\lambda_n(m + (i-k-1)h))$$
$$+ \delta_{m,1}\delta_{i,j}\delta_{k,l}\delta_{i,1}\delta_{k,1/h}(K_u - 1)V_u^2,$$

where the first part in the last equation is a consequence of Lemma A.1 given below. The last equation achieves the proof of (2.19). □

Lemma A.1. Let a, b, c, d be real numbers such that $a \leq b \leq c \leq d$. Then, for any $h > 0$,

$$\text{Cov}\left[\int_a^b \sigma_u^2 du, \int_c^d \sigma_u^2 du\right] = \sum_{n=1}^{p} \frac{a_n^2}{\lambda_n^2}[1 - \exp(-\lambda_n(b-a))]$$
$$\times [1 - \exp(-\lambda_n(d-c))]$$
$$\times \exp(-\lambda_n(c-b)). \tag{A.5}$$

Proof of Lemma A.1. We have

$$\text{Cov}\left[\int_a^b \sigma_u^2 du, \int_c^d \sigma_u^2 du\right]$$
$$= \sum_{1 \leq n,m \leq p} a_n a_m E\left[\int_a^b P_n(f_u) du \int_c^d P_m(f_u) du\right]$$
$$\equiv \sum_{1 \leq n,m \leq p} a_n a_m b_{n,m},$$

$$b_{n,m} = E\left[\int_a^b P_n(f_u)\mathrm{d}u \int_c^d E[P_m(f_u) \mid f_\tau, \tau \leq b]\mathrm{d}u\right]$$

$$= \int_a^b E[P_n(f_u)P_m(f_b)]\mathrm{d}u \int_c^d \exp(-\lambda_m(u-b))\mathrm{d}u$$

$$= \sum_{1 \leq n,m \leq p} a_n a_m \int_a^b \exp(-\lambda_n(b-u))\delta_{n,m}\mathrm{d}u$$

$$\times \frac{[1-\exp(-\lambda_m(d-c))]}{\lambda_m}\exp(-\lambda_m(c-b))$$

$$= \sum_{1 \leq n,m \leq p} a_n a_m \delta_{n,m} \frac{[1-\exp(-\lambda_n(b-a))]}{\lambda_n}$$

$$\times \frac{[1-\exp(-\lambda_m(d-c))]}{\lambda_m}\exp(-\lambda_n(c-b)),$$

i.e., (A.5).

In order to prove (2.20) note that the independence of the noise with the volatility and the no leverage assumption imply that

$$\mathrm{Cov}[\mathrm{IV}_t, r_{t-1+ih}^{(h)}r_{t-1+jh}^{(h)}]$$

$$= \mathrm{Cov}[\mathrm{IV}_t, r_{t-1+ih}^{*(h)}r_{t-1+jh}^{*(h)}] = E[\mathrm{IV}_t r_{t-1+ih}^{*(h)}r_{t-1+jh}^{*(h)}]$$

$$- E[\mathrm{IV}_t]E[r_{t-1+ih}^{*(h)}r_{t-1+jh}^{*(h)}]$$

$$= E[\mathrm{IV}_t E[r_{t-1+ih}^{*(h)}r_{t-1+jh}^{*(h)} \mid \sigma_\tau, \tau \leq t]] - a_0 \delta_{i,j} a_0 h$$

$$= \delta_{i,j}\left(E\left[\mathrm{IV}_t \int_{t-1+(i-1)h}^{t-1+ih} \sigma_u^2 \mathrm{d}u\right] - a_0^2 h\right)$$

$$= \delta_{i,j}\mathrm{Cov}\left[\left(\int_{t-1}^{t-1+(i-1)h} \sigma_u^2 \mathrm{d}u + \int_{t-1+(i-1)h}^{t-1+ih} \sigma_u^2 \mathrm{d}u\right.\right.$$

$$\left.\left. + \int_{t-1+ih}^{t} \sigma_u^2 \mathrm{d}u\right), \int_{t-1+(i-1)h}^{t-1+ih} \sigma_u^2 \mathrm{d}u\right]$$

$$= \delta_{i,j}\left(\sum_{n=1}^p \frac{a_n^2}{\lambda_n^2}[2 - \exp(-\lambda_n(i-1)h) - \exp(-\lambda_n(1-ih))]\right.$$

$$\left. \times [1 - \exp(-\lambda_n h)] + \mathrm{Var}\left[\int_{t-1+(i-1)h}^{t-1+ih} \sigma_u^2 \mathrm{d}u\right]\right), \quad (A.6)$$

where the last equality holds due to Lemma A.1. Eq. (15) in ABM (2004) implies

$$\mathrm{Var}\left[\int_{t-1+(i-1)h}^{t-1+ih} \sigma_u^2 \mathrm{d}u\right] = 2\sum_{n=1}^p \frac{a_n^2}{\lambda_n^2}[\exp(-\lambda_n h) + \lambda_n h - 1]. \quad (A.7)$$

By combining (A.6) and (A.7), one gets (2.20). Similar arguments lead to

$$\mathrm{Cov}[\mathrm{IV}_{t+1:t+m}, r_{t-1+ih}^{(h)}r_{t-1+jh}^{(h)}] = \mathrm{Cov}[\mathrm{IV}_{t+1:t+m}, r_{t-1+ih}^{*(h)}r_{t-1+jh}^{*(h)}]$$

$$= \delta_{i,j}\mathrm{Cov}\left[\mathrm{IV}_{t+1:t+m}, \int_{t-1+(i-1)h}^{t-1+ih} \sigma_u^2 \mathrm{d}u\right]$$

$$= \delta_{i,j}\left(\sum_{i=1}^p \frac{a_i^2}{\lambda_i^2}[1 - \exp(-\lambda_n h)][1 - \exp(-\lambda_n m)]\right.$$

$$\left. \times \exp(-\lambda_n(1-ih))\right), \quad (A.8)$$

where the last equality holds due to Lemma A.1. This achieves the proof of (2.21). $\square$

Proof of Proposition 4.2. The first line of (A.8) implies

$$\mathrm{Cov}[\mathrm{IV}_{t+1}, RM_t(h)] = \sum_{1 \leq i,j \leq 1/\underline{h}} q_{ij}\mathrm{Cov}[\mathrm{IV}_{t+1}, r_{t-1+ih}^{(h)}r_{t-1+jh}^{(h)}]$$

$$= \sum_{1 \leq i \leq 1/\underline{h}} q_{ii}\mathrm{Cov}\left[\mathrm{IV}_{t+1}, \int_{t-1+(i-1)h}^{t-1+ih} \sigma_u^2 \mathrm{d}u\right]$$

$$= \mathrm{Cov}[\mathrm{IV}_{t+1}, \mathrm{IV}_t]$$

under (4.27), which achieves the proof of (4.28) and Proposition 4.2. $\square$

Proof of Proposition 4.3. We have

$$\mathrm{Cov}[\overline{RM}_{t+m}(h), RM_t(h)] = \sum_{1 \leq i,j,k,l \leq 1/\underline{h}} \overline{q}_{ij}q_{kl}$$

$$\times \mathrm{Cov}[r_{t+m+ih}^{(h)}r_{t+m+jh}^{(h)}, r_{t+kh}^{(h)}r_{t+lh}^{(h)}]. \quad (A.9)$$

When $m > 1$, (A.9) combined with the first part of (2.19) leads to (4.32). When $m = 1$, (A.9) combined with the second and third parts of (2.19) leads to (4.31). This achieves the proof of Proposition 4.3. $\square$

References

Aït-Sahalia, Y., Hansen, L.P., Scheinkman, J., 2010. Operator Methods for Continuous-time Markov Models. In: Aït-Sahalia, Y., Hansen, L.P. (Eds.), Handbook of Financial Econometrics, North-Holland, pp. 1–66.

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

Aït-Sahalia, , Mykland, P.A., Zhang, L., 2005. How often to sample a continuous-time process in the presence of market microstructure noise. Review of Financial Studies 18, 351–416.

Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. International Economic Review 39, 885–905.

Andersen, T.G., Bollerslev, T., Christoffersen, P.F., Diebold, F.X., 2006. Volatility forecasting. In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), Handbook of Economic Forecasting. North-Holland, pp. 778–878.

Andersen, T.G., Bollerslev, T., Diebold, F.X., 2010. Parametric and Nonparametric Volatility Measurement. In: Aït-Sahalia, Y., Hansen, L.P. (Eds.), Handbook of Financial Econometrics, North-Holland, pp. 67–128.

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

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2001. The distribution of exchange rate volatility. Journal of the American Statistical Association 96, 42–55.

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

Andersen, T.G., Bollerslev, T., Lange, S., 1999. Forecasting financial market volatility: sample frequency vis-à-vis forecast horizon. Journal of Empirical Finance 6, 457–477.

Andersen, T.G., Bollerslev, T., Meddahi, N., 2004. Analytic evaluation of volatility forecasts. International Economic Review 45, 1079–1110.

Andersen, T.G., Bollerslev, T., Meddahi, N., 2005. Correcting the errors: volatility forecast evaluation using high-frequency data and realized volatilities. Econometrica 73, 279–296.

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

Areal, N.M.P.C., Taylor, S.J., 2002. The realized volatility of FTSE-100 futures prices. Journal of Futures Markets 22, 627–648.

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

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

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

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2008a. Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. Econometrica 76, 1481–1536.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2008b. Multivariate realised kernels: consistent positive semi-definite estimators of the covariance of equity prices with noise and non-synchronous trading. Working Paper. Oxford University.

Barndorff-Nielsen, O.E., Shephard, N., 2001. Non-Gaussian OU based models and some of their uses in financial economics, with discussion. Journal of the Royal Statistical Society, Series B 63, 167–241.

Barndorff-Nielsen, O.E., Shephard, N., 2002. Econometric analysis of realised volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society, Series B 64, 253–280.

Chen, X., Hansen, L.P., Scheinkman, J., 2009. Nonlinear principal components and long run implications of multivariate diffusions. The Annals of Statistics 37, 4279–4312.

Corradi, V., Distaso, W., Swanson, N.R., 2009a. Predictive density estimators for daily volatility based on the use of realized measures. Journal of Econometrics 150, 119–138.

Corradi, V., Distaso, W., Swanson, N.R., 2009b. Predictive inference for integrated volatility. Unpublished Manuscript. University of London.

Corsi, F., 2009. A simple approximate long memory model of realized volatility. Journal of Financial Econometrics 7, 174–196.

Deo, R., Hurvich, C., Lu, Y., 2006. Forecasting realized volatility using a long-memory stochastic volatility model: estimation, prediction and seasonal adjustment. Journal of Econometrics 131, 29–58.

Engle, R.F., Lee, G.G.J., 1999. A permanent and transitory component model for stock return volatility. In: Engle, R.F., White, H. (Eds.), Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive W.J. Granger. Oxford University Press, pp. 475–497.

Garcia, R., Lewis, M.A., Pastorello, S., Renault, E., 2011. Estimation of objective and risk-neutral distributions based on moments of integrated volatility. Journal of Econometrics 160 (1), 22–32.

Garcia, R., Meddahi, N., 2006. Comment on realized variance and market microstructure noise. Journal of Business and Economic Statistics 24, 184–191.

Ghysels, E., Santa-Clara, P., Valkanov, R., 2006. Predicting volatility: getting the most out of return data sampled at different frequencies. Journal of Econometrics 131, 59–96.

Ghysels, E., Sinko, A., 2006. Volatility forecasting and microstructure noise. Unpublished Manuscript. University of North Carolina.

Hansen, P.R., Lunde, A., 2006. Realized variance and market microstructure noise. Journal of Business and Economic Statistics 24, 127–161.

Hansen, L.P., Scheinkman, J., 1995. Back to the future: generating moment implications for continuous time Markov processes. Econometrica 63, 767–804.

Hull, J., White, A., 1987. The pricing of options on assets with stochastic volatilities. Journal of Finance XLII, 281–300.

Jacod, J., Protter, P., 1998. Asymptotic error distributions for the Euler method for stochastic differential equations. The Annals of Probability 26, 267–307.

Koopman, S.J., Jungbacker, B., Hol, E., 2005. Forecasting daily variability of the S&P 100 stock index using historical, realized and implied volatility measures. Journal of Empirical Finance 12, 445–475.

Martens, M., 2002. Measuring and forecasting S&P 500 index futures volatility using high-frequency data. Journal of Futures Markets 22, 497–518.

Meddahi, N., 2001. An eigenfunction approach for volatility modeling. CIRANO Working Paper. 2001s-70.

Meddahi, N., 2002a. A theoretical comparison between integrated and realized volatility. Journal of Applied Econometrics 17, 479–508.

Meddahi, N., 2002b. Moments of continuous time stochastic volatility models. Working Paper. Université de Montréal.

Merton, R.C., 1980. On estimating the expected return on the market: an exploratory investigation. Journal of Financial Economics 8, 323–361.

Patton, A.J., 2011. Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics 160 (1), 246–256.

Pong, S., Shackleton, M.B., Taylor, S.J., Xu, X., 2004. Forecasting currency volatility: a comparison of implied volatilities and AR(FI)MA models. Journal of Banking and Finance 28, 2541–2563.

Sun, X., 2006. Best quadratic unbiased estimators of integrated variance in the presence of market microstructure noise. Working Paper. University of California at San Diego.

Thomakos, D.D., Wang, T., 2003. Realized volatility in the futures market. Journal of Empirical Finance 10, 321–353.

Zhang, L., 2006. Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. Bernoulli 12, 1019–1043.

Zhang, L., Mykland, P.A., 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.

Zhou, B., 1996. High-frequency data and volatility in foreign exchange rates. Journal of Business and Economic Statistics 14, 45–52.

Zumbach, G., Corsi, F., Trapletti, A., 2002. Efficient estimation of volatility using high-frequency data. Unpublished Manuscript. Olsen & Associates, Zürich, Switzerland.