

Realized Volatility Forecasting and Market Microstructure Noise*

September 1, 2006

Torben G. Andersen*, Tim Bollerslev[†], and Nour Meddahi[‡]

Abstract

This paper extends the analytical results for reduced form realized volatility based forecasting in Andersen, Bollerslev and Meddahi (2004) to allow for market microstructure frictions and deviations from the basic semi-martingale assumption in the observed high-frequency returns underlying the realized measures. Our results build directly on the powerful eigenfunction representation of the general stochastic volatility class of models originally developed by Meddahi (2001). In addition to the traditional realized volatility measure and the role of the underlying sampling frequencies, we also explore the forecasting performance of several alternative volatility measures explicitly designed to mitigate the impact of the market microstructure noise. Our analysis of these alternative robust measures is facilitated by a simple unified quadratic form representation. Our results suggest that the detrimental impact of the noise in terms of the accuracy of the forecasts can be substantial, and that in empirically realistic situations the linear forecasts based on a simple-to-implement 'average' estimator defined by averaging across several sparsely sampled traditional realized volatility measures generally performs on par with best alternative robust measures.

Keywords: Volatility forecasting; high-frequency data; market microstructure noise; integrated volatility; realized volatility; robust volatility measures; eigenfunction stochastic volatility models.

JEL classification: C14, C22, C52, G14

*We are grateful to Peter R. Hansen and Neil Shephard for helpful discussions. We would also like to thank Bruno Feunou for excellent research assistance. The work of Andersen and Bollerslev was supported by a grant from the NSF to the NBER.

*Department of Finance, Kellogg School of Management, Northwestern University, Evanston, IL 60208, and NBER, USA, phone: 847-467-1285, e-mail: t-andersen@kellogg.northwestern.edu.

[†]Department of Economics, Duke University, Durham, NC 27708, and NBER, USA, phone: 919-660-1846, e-mail: boller@econ.duke.edu.

[‡]Tanaka Business School, Imperial College London.

1 Introduction

Over the last few decades, the notion that mean returns as well as return variances and covariances are time-varying has become widely accepted. It is also broadly recognized that direct inference regarding the conditional mean asset return without additional asset pricing assumptions is near impossible due to the amount of idiosyncratic return variation in most financial price series. There is simply too much noise in the data to infer the signal, i.e., the mean return, with any precision over even relatively long annual samples. However, the return variation and covariation are directly linked to the strength of the idiosyncratic return components and, to the extent these are dominant in the data, it should be feasible to measure their fluctuations over time with much better precision than for mean returns. This basic intuition is presented more formally in, e.g., Merton (1980). In particular, if the asset price and the associated return volatility process follow pure diffusions and price observations are available continuously, then it is trivial to determine the instantaneous return volatility at any point in time although it is impossible to pin down the corresponding instantaneous mean return.

In recent years, the increased availability of complete transaction and quote records for financial assets has spurred a literature seeking to exploit this additional source of information in estimating the current level of return variation. Inspired by the results discussed above, one may expect estimation of return variation to be straightforward given the abundance of data from active financial markets. However, practical implementation presents a number of significant challenges. First, even for very liquid assets, where lots of intraday price data are available, we do not have continuous observations of the underlying asset price but instead observe prices at frequent, yet intermittent, discrete points in time over the trading day. This induces an inevitable discretization error into estimates of current volatility. Second, and more importantly, the recorded prices do not reflect direct observations of a frictionless diffusive price process. Market prices are invariably quoted on a discrete price grid, there is a gap between buying and selling prices or quotes, i.e., a bid-ask spread, and different prices may be quoted by different market makers simultaneously due to heterogeneous beliefs, information and inventory positions. The latter set of complications is referred to jointly as market microstructure effects. Consequently, any observed intraday price does not correspond to a unique market price at a precise point in time but instead represents the underlying ideal theoretical price confounded by an error term reflecting the impact of market microstructure frictions, or "noise."

The early literature seeks to accommodate microstructure noise by sampling the prices sparsely relative to data availability. The guiding principle is the semi-martingale property of the underlying price process implied by a natural no-arbitrage condition. This property ensures that price increments over short intervals are approximately serially uncorrelated. As

a result, the realized volatility estimator, computed by cumulating intraday squared returns, provides a near unbiased realized return variation measure; see, e.g., Andersen, Bollerslev, Diebold and Labys (2000). Moreover, as emphasized in the early work by Andersen and Bollerslev (1998), Andersen, Bollerslev, Diebold and Labys (2001) and Barndorff-Nielsen and Shephard (2001), in the pure diffusion case with no microstructure noise present, this estimator is consistent for the underlying cumulative (integrated) return variance over the measurement horizon as the sampling interval shrinks towards zero. Thus, by sampling relatively sparsely, one may proceed as if microstructure noise is absent, yet the resulting estimator should in theory prove reasonably accurate. Importantly, the realized volatility approach represents a paradigm shift from estimation of instantaneous volatility to estimation of the (average) realization of the volatility process over a non-trivial interval, implicitly acknowledging that meaningful model-free measurement of instantaneous volatility simply is not feasible with actual transaction prices and price quotes that necessarily embody microstructure distortions.

The main disadvantage of the realized volatility estimator constructed from sparsely sampled data is that the discretization error can be substantial; see Barndorff-Nielsen and Shephard (2002a, 2002b) and Meddahi (2002a). This naturally raises the issue of how to potentially improve on the estimator by sampling more frequently, thus exploiting more of the available price data, without incurring excessive biases from the cumulative impact of microstructure noise. An entire literature has developed to address this topic, and it is now evident that improved realized return variation measures may be obtained through several alternative approaches. Important contributions include the first-order autocorrelation adjustment originally proposed by Zhou (1996), the notion of an optimal sampling frequency as discussed by Bandi and Russell (2005a, 2005b) and Aït-Sahalia, Mykland and Zhang (2005a), the average and two-scale estimator developed by Zhang, Mykland and Aït-Sahalia (2005), the multi-scale estimator of Zhang (2006), as well as the general kernel type estimator of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006); a recent insightful survey is provided by Hansen and Lunde (2006).

Meanwhile, another key issue concerns the use of the realized return variation measures in support of financial decision making. Real-time asset allocation, derivatives pricing and risk management is conducted given current expectations regarding the return volatility, or more generally the entire return distribution, over the planning horizon. The available realized volatility estimates at the decision point represent measures of the ex-post realized return variation over the preceding periods rather than the conditional variance over the planning horizon. Hence, even if the sequence of past realized volatility measures provide an accurate indication of recent return variation, it must be mapped into a predictor of the expected return variance over a future period, or some other quantity related to the

application at hand, to be directly useful. In contrast to the construction of optimal or robust ex-post realized return variation measures, this critical forecasting step is inevitably model dependent.

The realized volatility literature is unfortunately less developed on this important forecast dimension. A number of empirically oriented studies have compared the performance of forecasts based on relatively simple reduced form models for the realized variation measures to more traditional daily return based forecast procedures as well as market based predictions such as option-implied volatility forecasts; see, e.g., Andersen, Bollerslev, Diebold and Labys (2003), Deo, Hurvich and Lu (2006), Koopman, Jungbacker and Hol (2005), and Pong, Shackleton, Taylor and Xu (2004), among many others. It is generally found that the realized variation forecasts clearly dominate the standard stochastic volatility model forecasts based on daily data and they also appear to perform at least on par with the options based measures. In terms of a more analytic assessment of the potential of the procedures to improve performance the initial evidence was mainly generated from a handful of simulation based studies, which aside from being tied to a specific model, generally also ignored practical market microstructure features.¹

The model specific nature of these earlier studies was largely circumvented by the subsequent approach of Andersen, Bollerslev and Meddahi (henceforth ABM)(2004, 2005) who utilize the so-called eigenfunction stochastic volatility (ESV) framework of Meddahi (2001) in the development of analytic expressions for forecast performance spanning the entire class of stochastic volatility diffusions commonly used in the literature. In particular, the set-up delivers expressions for the optimal linear forecasts based on the history of past realized volatility measures and it allows for direct performance comparisons as the underlying sampling frequency of the intraday returns varies or the relevant measurement horizon changes.² It also facilitates analysis of the degree of artificial deterioration in forecast performance due to the reliance of (feasible) realized volatility measures as ex-post benchmarks for return variation in lieu of the true underlying integrated volatility. Nonetheless, the studies do not account for the fact that microstructure noise invariably also will impact practical measurements and forecast performance. In fact, there does not appear to be any direct way to analytically assess this issue within the existing literature.³

¹For example, Andersen and Bollerslev (1998) and Andersen, Bollerslev and Lange (1999) study the extent to which realized volatility forecasts improve on standard GARCH forecasts under correct model specification through the use of simulations from an underlying GARCH diffusion model, with their results indicating rather substantial potential gains.

²This same approach has also recently been adopted by Corradi, Distaso and Swanson (2005) in analyzing the entire predictive density for integrated volatility, as explored more generally in Corradi, Distaso and Swanson (2006).

³Bandi, Russell and Zhu (2006) provide empirical evidence that the use of properly chosen sampling frequencies in the construction of realized volatility measures may have important economic benefits in practical dynamic portfolio choice.

This paper provides an extension to the ABM studies by explicitly accounting for empirically realistic microstructure noise in the analytic derivation of realized volatility based forecasts. As noted, the literature on this topic is sparse and, apart from concurrent work by Aït-Sahalia and Mancini (2006) and Ghysels and Sinko (2006), basically non-existent. The latter papers may be seen as complementary to ours in the sense that they resort to numerical simulation methods or direct comparative empirical assessment in order to rank the various forecasting procedures and also study return generating processes and forecast not addressed here.⁴ For example, Aït-Sahalia and Mancini (2006) 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. Finally, a preliminary review of some results originally derived for the current project can also be found in Garcia and Meddahi (2006).

The remainder of the paper unfolds as follows. We begin in the next section by a brief discussion of the general theoretical framework, including the basic ESV model assumptions and realized volatility definitions, followed by the explicit analytical expressions for all of the requisite moments underlying our subsequent theoretical results. In Section 3 we present the optimal linear forecasting rules for the integrated volatility when the traditional realized volatility measure used in the construction of the forecasts are contaminated by market microstructure noise. We also quantify the deteriorating impact of the noise for the performance of the forecasts and directly analyze the notion of an "optimal" sampling frequency in plausible empirical situations. Moreover, we show how optimally combining the intraday squared returns in the construction of the integrated volatility forecasts does not materially improve upon the forecasts constructed from the realized volatilities based on equally weighted intraday squared returns. In Section 4 we demonstrate how many of the robust realized volatility measures explicitly designed to mitigate the impact of market microstructure noise may be conveniently expressed as a quadratic form in the highest possible intraday returns. This representation, in turn, facilitates direct derivation of analytical expressions for the corresponding optimal linear integrated volatility forecasts. We find that a relatively simple estimator constructed by averaging over several sparsely sampled traditional realized volatility measures generally performs among the best in a forecasting sense. Moreover, the differences among the competing realized volatility estimators can be quite substantial, thus directly highlighting the importance of considering the impact of the noise in actually observed prices. We also show that practically feasible realized volatility forecasting regressions based on the "wrong" realized volatility measure may falsely suggest almost no predictability, when in fact more than fifty percent of the day-to-day variation in the (latent) integrated volatility is predictable. In Section 5 we briefly discuss some potential

⁴We became aware of these projects some time after initiating our own work on these issues.

generalizations of the underlying model assumptions, while Section 6 concludes. All proofs are deferred to a 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, $\{S_t^*, 0 \leq t\}$, is continuous and determined by the stochastic differential equation (sde)

$$d \log(S_t^*) = \sigma_t dW_t, \quad (2.1)$$

where W_t denotes a standard Brownian motion, and the instantaneous, or spot, volatility process σ_t is predictable with continuous sample path. Initially we assume that the σ_t and W_t processes are uncorrelated, ruling out so-called leverage effects, but we discuss this extension along with the impact of including a drift term in the specification in Section 5 below. For notational simplicity we henceforth refer to the unit time interval as a day.

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,

$$IV_{t+1} \equiv \int_t^{t+1} \sigma_\tau^2 d\tau, \quad (2.2)$$

and the corresponding multi-period measure,

$$IV_{t+1:t+m} = \sum_{j=1}^m IV_{t+j}, \quad (2.3)$$

where m is an integer. As discussed at length in Andersen, Bollerslev, and Diebold (2006) and Andersen, Bollerslev, Christoffersen and Diebold (2006), the integrated volatility provides *the* natural measure of the ex-post return variability.⁵

The integrated volatility is, of course, not directly observable. However, as highlighted in the work by Andersen and Bollerslev (1998), Andersen, Bollerslev, Diebold and Labys (2001, 2003), Barndorff-Nielsen and Shephard (2001, 2003), Meddahi (2002a), the realized volatilities defined by the summation of high-frequency squared returns from the diffusion in (2.1),

$$RV_t^*(h) \equiv \sum_{i=1}^{1/h} r_{t-1+ih}^{*(h)2}, \quad (2.4)$$

⁵The integrated volatility also figures prominently in the option pricing literature; see, e.g., Hull and White (1987) and Garcia, Lewis, Pastorello and Renault (2004).

where

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

and $1/h$ is assumed to be an integer, may in theory be used in the construction of arbitrarily accurate estimates of IV_t for sufficiently small values of h . That is, $RV_t^*(h)$ is uniformly consistent for IV_t as $h \rightarrow 0$, or the sampling frequency of the intraday returns underlying the realized volatility measure goes to infinity. Moreover, ABM (2004) have formally shown that simple autoregressive models for $RV_t^*(h)$ based on high values of h , or a large number of high-frequency returns, may be used in the construction of simple-to-implement and, for many of the stochastic volatility models employed in the literature, remarkably close to efficient forecasts for IV_{t+1} and $IV_{t+1:t+m}$.⁶

In practice, of course, the prices of even the most liquid assets are invariably affected by a host of market microstructure frictions, and as such do not adhere to the stylized diffusion model in equation (2.1) when observed at ultra-high frequencies. In response to this issue the studies above instead advocate using a moderately high but not ultra-high sampling frequency, so that the model in (2.1) may still be considered an adequate approximation to the observed price process (see, e.g., the discussion of the so-called volatility signature plot in Andersen, Bollerslev, Diebold and Labys, 2000, used in informally determining an appropriate value of h). Meanwhile, to better mimic actually observed prices when analyzing the properties of realized volatility type measures based on increasingly finer sampled returns, a number of recent studies have argued for the relevance of explicitly including an additional term in the price process to properly reflect the impact of the market microstructure "noise" (Hansen and Lunde, 2006, provides an insightful recent survey). Following the common setup in this literature, we will here assume that the actually observed price process, $\{S_t, 0 \leq t\}$, is equal to the semi-martingale price process defined in (2.1) plus a noise component, u_t ,

$$\log(S_t) = \log(S_t^*) + u_t, \quad (2.6)$$

where u_t is *i.i.d.* with mean zero, variance V_u , and kurtosis $K_u = E[u_t^4]/V_u^2$. In the numerical calculations reported on below we will focus on $K_u = 3$, corresponding to a normally distributed noise term, but our theoretical results are general and allow for any value of K_u . Also, even though our initial derivations and corresponding numerical calculations are based on the simplifying assumption of an *i.i.d.* noise component, we will briefly discuss the implications of relaxing that assumption and allowing for correlated noise in Section 5 below.

⁶These theoretical results directly corroborate the empirical findings reported in Andersen, Bollerslev, Diebold and Labys (2003), Areal and Taylor (2002), Corsi (2003), Deo, Hurvich and Lu (2006), Koopman, Jungbacker and Hol (2005), Martens (2002), Pong, Shackleton, Taylor and Xu (2004), and Thomakos and Wang (2003), among many others, involving the estimation of relatively simple reduced form time series forecasting models for various realized volatility series.

Assuming that S_t , but not S_t^* , is observable, the observed h -period intraday returns is correspondingly defined by,

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

Of course, the noise contaminated returns are directly related to the non-contaminated, or ideal, returns in (2.5) by the equation,

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

where

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

Under the assumption that u_t is *i.i.d.*, the noise therefore induces an MA(1) error structure in the observed returns. Moreover, for very large values of h the variance of the noise term, $e_t^{(h)}$, will tend to dominate the variance of the signal, or the true latent return $r_t^{*(h)}$. In fact, as shown by Bandi and Russell (2005a, 2005b) and Zhang, Mykland and Aït-Sahalia (2005) the practically feasible realized volatility measure based on the contaminated high-frequency returns,

$$RV_t(h) \equiv \sum_{i=1}^{1/h} r_{t-1+ih}^{(h)2} \quad (2.10)$$

is formally inconsistent for IV_t and diverges to infinity for $h \rightarrow 0$. Of course, this does not mean that $RV_t(h)$ can not be used in meaningfully forecasting IV_{t+1} and $IV_{t+1:t+m}$ for moderately large values of h . Indeed, as we document below, by choosing the value of h to appropriately balance the impact of the noise and the signal in empirically realistic situations, remarkably accurate volatility forecasts based on simple autoregressive models for $RV_t(h)$ are still feasible. However, as discussed further in Section 4, a number of alternative robust realized volatility measures explicitly designed to account for the high-frequency noise have recently been proposed in the literature. As part of our analysis we therefore also compare and contrast the performance of reduced form forecasting models for these alternative measures to the forecasts based on the traditional $RV_t(h)$ measure in equation (2.10).

2.2 Eigenfunction Stochastic Volatility Models

To facilitate the derivation of our theoretical results we follow ABM (2004) in assuming that the spot volatility process is a member of the Eigenfunction Stochastic Volatility (ESV) class of models introduced by Meddahi (2001). This is a very general class of models that includes most continuous-time diffusive stochastic volatility models in the existing literature.

To illustrate, consider the situation in which the volatility process is driven by a single (latent) state variable.⁷ The corresponding one-factor ESV representation then takes the generic form,

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

where the integer p may be infinite, and the latent state variable f_t is characterized by the process,

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

where the W_t^f Brownian motion is independent of the W_t Brownian motion driving the price in equation (2.1). Further, the a_n coefficients in the sum are real numbers and the $P_n(f_t)$'s denote the eigenfunctions of the infinitesimal generator associated with f_t .⁸ The power and convenience of the ESV representation stems from the fact that the eigenfunctions are orthogonal and centered at zero,

$$E[P_n(f_t)P_j(f_t)] = 0 \quad E[P_n(f_t)] = 0, \quad (2.13)$$

and follow first-order autoregressive processes,

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

where $(-\lambda_n)$ denote the corresponding eigenvalues. These simplifying features jointly render the calculation of analytical multi-step forecasts for σ_t^2 and the moments of discretely sampled returns, $r_t^{(h)}$, from the model defined by equations (2.1), (2.8), (2.11) and (2.12), feasible. The following proposition provides the relevant analytical expressions that we need in our subsequent analysis.

Proposition 2.1 *Let the discrete-time noise contaminated and non-contaminated returns, $r_t^{(h)}$ and $r_t^{*(h)}$, respectively, be determined by an ESV model and equation (2.8), with the corresponding one- and m -period integrated volatilities, IV_t and $IV_{t+1:t+m}$, defined by equations (2.2) and (2.3), respectively. Then for positive integers $i \geq j \geq k \geq l$, $m \geq 1$, and $h > 0$,*

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

⁷The 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, Hansen and Scheinkman (2005) for a general approach to eigenfunction modeling for multivariate Markov processes.

⁸For a more detailed discussion of the properties of infinitesimal generators see, e.g., Hansen and Scheinkman (1995) and Ait-Sahalia, Hansen and Scheinkman (2006).

$$E[r_{t+ih}^{(h)2}] = Var[r_{t+ih}^{(h)}] = Var[r_{t+ih}^{*(h)}] + 2V_u = a_0h + 2V_u, \quad (2.16)$$

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

$$E[r_{t+ih}^{(h)}r_{t+jh}^{(h)}r_{t+kh}^{(h)}r_{t+lh}^{(h)}] = \quad (2.18)$$

$$\begin{aligned} & 3a_0^2h^2 + 6 \sum_{n=1}^p \frac{a_n^2}{\lambda_n^2} [-1 + \lambda_n h + \exp(-\lambda_n h)] + 2V_u^2(K_u + 3) + 12a_0V_uh \text{ if } i = j = k = l, \\ & - V_u^2(K_u + 3) - 3a_0V_uh \text{ if } i = j = k = l + 1 \text{ or } i = j + 1 = k + 1 = l + 1, \\ & a_0^2h^2 + \sum_{n=1}^p \frac{a_n^2}{\lambda_n^2} [1 - \exp(-\lambda_n h)]^2 + V_u^2(K_u + 3) + 4a_0V_uh \text{ if } i = j = k + 1 = l + 1, \\ & a_0^2h^2 + \sum_{n=1}^p \frac{a_n^2}{\lambda_n^2} [1 - \exp(-\lambda_n h)]^2 \exp(-\lambda_n(i - k - 1)h) + 4V_u^2 + 4a_0V_uh \text{ if } i = j > k + 1, k = l, \\ & 2V_u^2 \text{ if } i = j + 1, j = k = l + 1, \\ & - 2V_u^2 - a_0V_uh \text{ if } i = j \geq k + 1, k = l + 1, \text{ or } i = j + 1, j \geq k + 1, k = l, \\ & V_u^2 \text{ if } i = j + 1, j \geq k + 1, k = l + 1, \\ & 0 \text{ otherwise.} \end{aligned}$$

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

$$\begin{aligned} & \sum_{n=1}^p \frac{a_n^2}{\lambda_n^2} (1 - \exp(\lambda_n h))^2 \exp(-\lambda_n(m + (i - k - 1)h)) \text{ if } m \geq 2, i = j, k = l, \\ & \sum_{n=1}^p \frac{a_n^2}{\lambda_n^2} (1 - \exp(\lambda_n h))^2 \exp(-\lambda_n(1 + (i - k - 1)h)) \text{ if } m = 1, i = j, k = l, i \neq 1, k \neq 1/h, \\ & \sum_{n=1}^p \frac{a_n^2}{\lambda_n^2} (1 - \exp(\lambda_n h))^2 + (K_u - 1)V_u^2 \text{ if } m = 1, i = j, k = l, i = 1, k = 1/h, \\ & 0 \text{ otherwise.} \end{aligned} \quad (2.19)$$

$$E[IV_t r_{t-1+ih}^{(h)} r_{t-1+jh}^{(h)}] =$$

$$\begin{aligned} & a_0^2h + 2a_0V_u + 2 \sum_{n=1}^p \frac{a_n^2}{\lambda_n^2} (\exp(-\lambda_n h) + \lambda_n h - 1) \\ & + \sum_{n=1}^p \frac{a_n^2}{\lambda_n^2} (2 - \exp(-\lambda_n(i - 1)h) - \exp(-\lambda_n(1 - ih))) (1 - \exp(-\lambda_n h)) \text{ if } i = j, \quad (2.20) \\ & - a_0V_u \text{ if } |i - j| = 1, \\ & 0 \text{ otherwise.} \end{aligned}$$

$$\begin{aligned}
E[IV_{t+1:t+m} r_{t-1+ih}^{(h)} r_{t-1+jh}^{(h)}] = & \\
& a_0^2 hm + 2a_0 m V_u \\
& + \sum_{n=1}^p \frac{a_n^2}{\lambda_n^2} (1 - \exp(-\lambda_n h))(1 - \exp(-\lambda_n m) \exp(-\lambda_n(1 - ih))) \quad \text{if } i = j, \\
& - a_0 m V_u \quad \text{if } |i - j| = 1, \\
& 0 \quad \text{otherwise.}
\end{aligned} \tag{2.21}$$

Proof: See Appendix II.

In the numerical calculations reported on below, we focus on the same three models previously analyzed in ABM (2004, 2005), namely a GARCH diffusion model, a two-factor affine model, and a log-normal diffusion. These particular models are fairly representative of the models and parameter values entertained in the broader literature. The corresponding ESV representations for each of the models are given in Appendix I. Here, we simply state the there models in the more familiar sde form.

Model M1 - GARCH Diffusion The instantaneous volatility in the GARCH diffusion model is defined by the process,

$$d\sigma_t^2 = \kappa(\theta - \sigma_t^2)dt + \sigma\sigma_t^2 dW_t^{(2)}, \tag{2.22}$$

where $\kappa = 0.035$, $\theta = 0.636$, and $\psi = 0.296$.

Model M2 - Two-Factor Affine The instantaneous volatility in the two-factor affine model is given by,

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

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.

Model M3 - Log-Normal Diffusion The instantaneous volatility in the log-normal diffusion follows the process,

$$d \log(\sigma_t^2) = \kappa[\theta - \log(\sigma_t^2)]dt + \sigma dW_t^{(2)}, \tag{2.24}$$

where $\kappa = 0.0136$, $\theta = -0.8382$, and $\sigma = 0.1148$.

This completes our discussion of the general ESV model framework that underlie our main results. In the next section, we move on to discuss the properties of integrated volatility forecasts based on the traditional realized volatility measure in the presence of noise.

3 Traditional Realized Volatility Based Forecasts

3.1 Optimal Linear Forecasts

ABM (2004) provide a detailed assessment of integrated volatility forecasts constructed from the linear regression of IV_{t+1} and $IV_{t+1:t+m}$ on $RV_t^*(h)$ and a constant, as well as longer regressions including additional lagged values of $RV_t^*(h)$. However, as discussed in the previous section, the stylized semi-martingale price process in equation (2.1) breaks down when prices are sampled at ultra-high frequencies, and the modified price process in (2.6) arguably provides a more accurate description of real-world prices and finely sampled returns. As such, theoretical results based on $RV_t(h)$ in place of $RV_t^*(h)$ for large values of h , should provide better guidance for practical empirical work and afford a direct assessment of the deteriorating impact of the market microstructure noise in terms of the accuracy of the forecasts.

In order to derive analytical expressions for the corresponding linear forecasts of IV_{t+1} and $IV_{t+1:t+m}$ based on $RV_t(h)$ and lagged values of $RV_t(h)$, we need to calculate $Cov[IV_{t+1}, RV_{t-l}(h)]$, $Var[RV_t(h)]$, and $Cov[RV_{t+1}(h), RV_{t-l}(h)]$, for $l \geq 0$. To do so, note that from equation (2.8) the contaminated and the ideal realized volatility measures are directly related by the equation,

$$RV_t(h) = RV_t^*(h) + \sum_{i=1}^{1/h} e_{t-1+ih}^{(h)2} + 2 \sum_{i=1}^{1/h} r_{t-1+ih}^{*(h)} e_{t-1+ih}^{(h)}. \quad (3.1)$$

Utilizing this decomposition the following general set of results for any pair of price processes defined by equations (2.1) and (2.6) follow fairly easily.

Proposition 3.1 *Let the discrete-time noise contaminated and non-contaminated returns, $r_t^{(h)}$ and $r_t^{*(h)}$ respectively, be determined by the diffusion in (2.1) and equation (2.8), with the corresponding realized and integrated volatilities, $RV_t^*(h)$, $RV_t(h)$, IV_t and $IV_{t+1:t+m}$, defined by equations (2.4), (2.10), (2.2) and (2.3), respectively. Then for integers $m \geq 1$ and $l \geq 0$, and $h > 0$,*

$$Cov[IV_{t+1:t+m}, RV_{t-l}(h)] = Cov[IV_{t+1:t+m}, RV_{t-l}^*(h)] = Cov[IV_{t+1:t+m}, IV_{t-l}], \quad (3.2)$$

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

$$Cov[RV_{t+1}(h), RV_t(h)] = Cov[RV_{t+1}^*(h), RV_t^*(h)] + (K_u - 1)V_u^2, \quad (3.4)$$

$$Cov[RV_{t+1}(h), RV_{t-l}(h)] = Cov[RV_{t+1}^*(h), RV_{t-l}^*(h)]. \quad (3.5)$$

Proof: See Appendix II.

This proposition expresses the variances and covariances for $RV_t(h)$ as explicit functions of the corresponding counterparts for the ideal $RV_t^*(h)$ measure along with the variance and kurtosis of the noise. Analytical expressions for the latter quantities valid across the ESV class of stochastic volatility models, implicitly defined by equation (2.11), are derived in ABM (2004). Specifically, adapting their notation, we have the following results,

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

$$Cov(IV_{t+1:t+m}, IV_{t-l}) = \sum_{n=1}^p a_n^2 \frac{[1 - \exp(-\lambda_n)][1 - \exp(-\lambda_n m)]}{\lambda_n^2} \exp(-\lambda_n l), \quad (3.7)$$

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

and

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

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

It is, of course, impossible to quantify the detrimental impact from market microstructure noise on the precision of forecasts of future return variation based on the history of realized volatility measures in general, as the appropriate loss function depends on the economic application that the volatility forecasts is to be used for. Instead, we follow the literature in focusing on the coefficient-of-determination, or R^2 , from the regression of the future integrated variance on a constant and the associated forecast variables. This implicitly corresponds to the use of mean-squared-forecast-error (MSE) criterion for the unconditional bias-corrected return variation forecast. This is a common benchmark for performance evaluation and it allows for direct comparison with prior results in the literature which have been assessed within a similar setting.⁹

In particular, following ABM (2004), 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,

$$R^2(IV_{t+1}, RV_t(h), l) = \frac{C(IV_{t+1}, RV_t(h), l)^\top (M(RV_t(h), l))^{-1} C(IV_{t+1}, RV_t(h), l)}{Var[IV_t]}. \quad (3.10)$$

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,

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

⁹Patton (2006) 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.

and,

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

respectively. The corresponding R^2 for the longer-horizon integrated volatility forecasts is simply obtained by replacing IV_{t+1} with $IV_{t+1:t+m}$ in the formulas immediately above. We next turn to our discussion of the resulting numerical R^2 's for the three specific ESV models discussed in Section 2 when contaminated by different levels of noise.

3.2 Quantifying the Impact of Market Microstructure Noise

The impact of the microstructure noise is, of course, directly related to the magnitude of the variation in the noise relative to the true daily return variation. In our setting, this factor is conveniently captured by the noise-to-signal ratio, or $\lambda \equiv V_u/E[IV_t]$. Hansen and Lunde (2006) estimate this factor across thirty active U.S. stocks for the year 2000 and find typical values to be around 0.1%, with most being slightly less. They also note that the magnitude of the noise has generally decreased since then and now is substantially lower for many actively traded stocks. Consequently, we use 0.1% as a reference point for a realistic, and for liquid assets perhaps even inflated, value for the relative size of the noise component. We also explore the impact of significantly higher levels of microstructure noise by reporting results for noise-to-signal ratios of 0.5% and 1.0%.

Table 1 to be inserted here.

Table 1 summarizes the findings in terms of the population R^2 in equation (3.10) from the regression of future integrated volatility on the various realized measures across different forecast horizons, data generating processes (models), levels of microstructure noise, and sampling frequencies for the intraday returns. As a reference point, we also include in the first row the R^2 's for the optimal (infeasible) forecasts based on the exact value of the (latent) volatility state variable(s) in the ESV models. Similarly, the next two rows report the (infeasible) forecasts based on the past daily (latent) integrated volatility and the forecasts exploiting an additional four lags, or a week, of the integrated volatility. The next eleven rows then give the R^2 's for different realized volatility based forecasts under the assumption of no microstructure noise, or $\lambda = 0$, and sampling frequencies ranging from $h = 1/1444$ to $h = 1$, corresponding to 1-minute to daily return observations in 24-hour market, and we shall refer to them as such in our subsequent discussion. However, the highest $h = 1/1444$ frequency may equally well be interpreted as arising from 15-second sampling over a 6-hour trading day. All of these reference scenarios were previously analyzed in ABM (2004) and

represent a small subset of the findings discussed herein.¹⁰

From the first row, it is evident that Models 1 and 3 imply an inherently large degree of predictability in return variation, while this is slightly less so for Model 2. The latter embodies a second, less persistent, factor which limits the overall amount of serial correlation in the return volatility process. From the second and third rows we see that there is only a small loss of predictability associated with forecasts based on the last day's integrated volatility rather than the instantaneous volatility, and exploiting the last week's worth of daily integrated volatilities is at best only marginally helpful. Again, the decline in predictive power is more pronounced for the less persistent process associated with model 2. Typically, adding additional lags has no discernable impact on the results based on intraday returns so we limit ourselves to the weekly window.

Moving on to the realized volatility based forecasts, albeit only for the ideal case without any noise, associated with rows four to fourteen, we see that the drop in predictive power is relatively small for the forecasts based on measures constructed from 1- and 5-minute returns. Moreover, as we move down to lower sampling frequencies for which the return variation measures are less precise, the addition of lagged daily realized volatility measures start to become more valuable. The results reported for twenty daily lags (19 additional) in row fourteen also closely mimic what would be achieved by a well-specified GARCH model, as detailed in ABM (2004).

Turning to the new results concerning the forecasts based on realized volatilities constructed from noisy return observations, we first observe that the degradation in performance is relatively mild for the realistic case with $\lambda = 0.1\%$. However, as the noise variance is increased in the bottom part of the table, the performance deteriorates more sharply, even if the addition of lagged values of the realized volatility measure now is critical in boosting the predictive power. A second general finding is that, as anticipated, it is not optimal to compute the realized volatility measures from returns corresponding to the very highest frequencies. At the moderate noise level of $\lambda = 0.1\%$, the performance is better for 5-minute rather than 1-minute sampling, and as λ grows further, sampling at the 15- and 30-minute levels, respectively, produce the highest coherence between forecasts and future realizations. This is, of course, the qualitative results we would expect from theory when the noise increasingly dominates the sampling variability as the main source of variation in the realized measures. Further evidence along these lines is obtained by comparing the relative decline in predictability from the $\lambda = 0.1\%$ to the $\lambda = 1.0\%$ scenarios for an underlying 5-minute ($h = 1/288$) versus 30-minute ($h = 1/48$) sampling frequency. One finds a drop in the R^2 from moderate to large noise and $h = 1/288$ at the one-day forecast horizon in Model

¹⁰Specifically, the figures reported in the first fourteen rows of Table 1 were extracted from Tables 1 through 6 in ABM (2004).

1 of about 91% to 47% (92% to 72% if lags are exploited) compared to a corresponding drop for $h = 1/48$ of about 82% to 65% (88% to 81% with lags). Third, we note that the importance of exploiting the information in lagged daily realized volatility measures increases sharply with the level of the noise. Even for $\lambda = 0.1\%$ the measures based on 30-minute cumulative squared returns are quite competitive with those based on 5-minute sampling once the lagged realized volatility measures are exploited. In fact, for the higher noise levels, the 30-minute based measures dominate the 5-minute based ones for all our scenarios. Hence, within the class of linear realized volatility based forecast procedures, the use of an underlying 30-minute return horizon appears to provide a robust and reasonably efficient choice as long as the series of past daily measures are incorporated into the construction of the forecasts.

3.3 Optimal Sampling Frequency

The results in the preceding section immediately point to the notion of an optimal sampling frequency for the high-frequency returns underlying $RV_t(h)$, in the sense of maximizing the R^2 associated with the linear forecasting regressions, or minimizing the MSE associated with the forecast errors. In this section we consider two different ways of defining an approximately optimal (unconditional) value of h . For concreteness, we focus our discussion on the one-step-ahead forecasts, but the same results are readily extended to a multi-period setting, as further illustrated in our numerical calculations.

Our first approach is based on the results of Bandi and Russell (2005a, 2005b) and Zhang, Mykland and Ait-Sahalia (2005). In particular, as showed therein, the optimal sampling frequency in the sense of minimizing the MSE of $RV_t(h)$ *conditional* on the volatility sample path is approximately given by,

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

where the integrated quarticity is defined by,

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

However, instead of attempting to estimate the optimal frequency on a period-by-period basis, we follow Bandi and Russell (2005a) in replacing the hard-to-estimate one-period integrated quarticity by its unconditional expectation. In particular, we consider the following unconditional counterpart to h_t^* ,

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

which is relatively easy to estimate and implement in practice.

In addition, we also consider the sampling frequency which minimizes the unconditional variance of $RV_t(h)$. To motivate this alternative choice of h , consider the R^2 from the regression of IV_{t+1} on a constant and $RV_t(h)$,

$$R^2 = \frac{Cov[IV_{t+1}, RV_t(h)]^2}{Var[IV_t]Var[RV_t(h)]} = \frac{Cov[IV_{t+1}, IV_t]^2}{Var[IV_t]Var[RV_t(h)]} \quad (3.16)$$

where the last equality follows from the result in Proposition 3.1. Consequently, maximizing this R^2 is equivalent to minimizing the variance of $Var[RV_t(h)]$, as first noted out by Ghysels and Sinko (2006). To minimize this variance, we follow Jacod and Protter (1998), Barndorff-Nielsen and Shephard (2002), and Meddahi (2002a), in approximating the variance of the corresponding non-contaminated realized volatility measure by,

$$Var[RV_t^*(h)] \approx Var[IV_t] + 2hE[IQ_t]. \quad (3.17)$$

Substituting this expression into the equation for $Var[RV_t(h)]$ in equation (3.3) immediately yields,

$$Var[RV_t^*(h)] \approx Var[IV_t] + 2hE[IQ_t] + 2V_u^2 \left(\frac{2K_u}{h} - K_u + 1 + 4\frac{E[\sigma_t^2]}{V_u} \right). \quad (3.18)$$

Minimizing this expression with respect to h result in our second approximate optimal sampling frequency,

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

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 large values of h , to assess V_u and K_u , along with the use of lower frequency returns to estimate $E[IQ_t]$; see Bandi and Russell (2005a, 2005b) for further discussion and empirical analysis along these lines.

Table 2 to be inserted here.

Table 2 reports the approximate optimal sampling frequencies, as represented by h_1 and h_2 , for the same scenarios analyzed in Table 1, along with the resulting population R^2 's. Given that the frequency h_2 by construction optimizes this value, it is not surprising that the associated forecasts uniformly perform better than those derived from the values of h_1 . However, the size of the discrepancy is striking given the implicit claim in the existing literature that h_1 should be a sensible guide for practical applications. In some cases, the R^2 increases by over 25% and there are always a few percent to be gained by adhering to the choice of h_2 rather than h_1 . The reason for these findings is readily apparent as h_2 invariably

involves sampling much more frequently than h_1 . This feature reflects the pronounced right skew in the distribution of the integrated quarticity which is the main determinant for the discretization error of the daily realized volatility measure. In particular, a high (estimated) value of the daily integrated quarticity is associated with a higher optimal sampling frequency for that day. Given the right skewed distribution, the average optimal sampling frequency is thus much lower than the optimal frequency for the highest realization of daily integrated quarticity, but the latter days are associated with much larger measurement errors than experienced for the average day. Hence, simply averaging the optimal choice of sampling frequency across trading days, as in the derivation of h_1 , ignores the disproportional losses suffered on the most volatile days. In contrast, h_2 directly minimizes the average squared error and thus adjusts the choice of sampling frequency to accommodate the excessive errors incurred on the more extreme days. Direct inspection of the formulas for (the inverse of) h_1 and h_2 also shows that the latter implies a sampling frequency that is more responsive to the size of the (unconditional) integrated quarticity than the former.

Of course, if the loss of sticking with a constant sampling frequency is large, one may be tempted to develop a strategy involving the choice of a time-varying sampling frequency depending on some initial estimate of the daily integrated quarticity. However, direct comparison of the performance of the forecasts associated with h_2 and moderate noise of 0.1% in Table 2 and the forecasts derived from realized volatility in the absence of microstructure noise in rows four through seven in Table 1 shows that the loss, in fact, is quite small. In summary, for the loss function and models considered here, it is generally more important to pin down a sensible sampling frequency than it is to vary the size of the intraday return interval from day to day in response to the perceived variation in the degree of precision of the realized volatility measure. This is obviously a comforting finding from a practical perspective.

These results are further corroborated by Figure 1, which shows the R^2 for the one-step-ahead forecasts for IV_{t+1} based on $RV_t(h)$ for different values of h , including h_1 and h_2 , for each of the two reference models. For the lowest, yet realistic, $\lambda = 0.1\%$ noise level in Table 2 the results for h_2 prescribe sampling somewhere between every five and one minute. This same result is readily apparent from the corresponding R^2 lines for the 5- and 1-minute frequencies, or $h = 1/288$ and $h = 1/1440$, in Figure 1, which are both close to the upper most envelope for values of the noise in that general range. This, of course, is also entirely consistent with the results reported in Table 1. Meanwhile, as the noise level increases the preferred sampling frequency obviously also decreases, and the detrimental impact of sampling too often at say the 1-minute level for values of λ in excess of say 0.25% is quite striking. Also, directly in line with the more detailed discussion above, the R^2 's for h_1 are systematically below those for h_2 , and often quite considerably so.

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 microstructure noise is present, the basic measures become inconsistent even if the sparse sampling estimators only suffer from minor finite sample biases. Nonetheless, these inconsistent measures still provide a sensible basis for predicting the future return variation through the forecast regressions discussed in the preceding sections, which automatically correct for any systematic (unconditional) bias through the inclusion of a constant term. In terms of forecasting, the issue for the regressors included on the right-hand-side is not the extent of any bias but rather the ability to capture relevant variation in current realized volatility, which then translates into improved predictive performance. These considerations suggest that we may want to further 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 observed intraday squared returns. Hence, 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

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

While this regression is difficult to implement in practice because of the large number of parameters, $1 + 1/h$, we can readily compute its population counterpart within the ESV setting from the results in Proposition 2.1. The corresponding numerical results are presented in Table 3. Given the similarity in findings for Models 1 and 3 throughout and the arguably excessive noise level associated with $\lambda = 1\%$ we have, for the sake of brevity and clarity, eliminated Model 3 as well as the $\lambda = 1\%$ scenario from this and all future tables.

Table 3 to be inserted here.

Comparing the results to the ones in the previous tables, the very small gains obtained by an optimal weighting scheme for the intraday squared returns are striking. Even if the improvements are slightly larger for Model 2 than the other models, they remain sufficiently small that they surely will be negated, in practice, through the need to estimate the many weighting scheme coefficients a priori. In short, the realized volatility estimator is not overly restrictive in terms of the weighting scheme applied to the series of intraday squared returns. Of course, this does not imply that alternative regressors constructed from the series of intraday returns may not outperform the realized volatility estimators in this regard. The above regression only allows for a linear weighting of the intraday squared returns, whereas

many microstructure robust estimators involve more complex combinations of the leads and lags of the intraday returns.¹¹ In the following section, we provide a more detailed investigation of these approaches in terms of their forecast potential.

4 Robust Realized Volatility Based Forecasts

In this section we investigate the extent to which simple reduced form forecasting models based on alternative noise robust realized variation measures improve upon the forecasts constructed on the basis of the traditional realized volatility measures as discussed above. In particular, we will consider the average and two-scale estimators of Zhang, Mykland and Aït-Sahalia (2005), the first-order autocovariance adjusted estimator of Zhou (1996), and the general kernel type estimator of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006).

4.1 Quadratic Form Representation

We begin our analysis by developing a unified quadratic form representation for each of the different estimators. In particular, let \underline{h} denote the shortest possible intra-day return interval and assume, without loss of generality, that $1/\underline{h}$ is an integer. As before, let $1/h$ denote the actual number of equally spaced intraday returns used in the construction of a specific (sparsely sampled) realized volatility estimator. It is convenient to write each such measure as a quadratic function of the $1/\underline{h} \times 1$ vector of the highest frequency returns. That is,

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

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

$$VR_t(\underline{h}) = (r_{t-1+\underline{h}}^{(h)}, r_{t-1+2\underline{h}}^{(h)}, \dots, r_t^{(h)})^\top. \quad (4.2)$$

In order to study the interaction of a particular realized volatility measure with other realized volatility measures as well as current and future values of the integrated variance, we need analytical expressions for the corresponding means, variances and covariances. The following proposition provides these quantities.

Proposition 4.1 *Let the discrete-time noise contaminated returns be determined by the diffusion in (2.1) and the relationship in equation (2.8). Let $\overline{RM}_t(h)$ and $RM_t(h)$ denote*

¹¹Likewise, the so-called MIDAS scheme of Ghysels, Santa-Clara and Valkanov (2006) utilizes nonlinear functions of past intraday absolute returns in the construction of volatility forecasts. These quantities are fully divorced from the notion of consistent or unbiased return variation measures, and as such also fall outside the ESV analytical framework and techniques explored in the present paper.

two realized volatility measures as defined in equation (4.1) with corresponding quadratic form weights q_{ij} and \bar{q}_{ij} , respectively. Also, let the integrated volatilities, IV_t and $IV_{t+1:t+m}$, be defined by equations (2.2) and (2.3), respectively. Then,

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

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

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

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

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

Proof: Follows directly from the quadratic form representation.

Closed form expressions for all of the expectations on the right-hand-sides of the expressions in the proposition are immediately available from the results for the general ESV class of models previously presented in Proposition 2.1.

In the following section we explicitly define the $(1/\underline{h} \times 1/\underline{h})$ matrix Q for each of the alternative realized volatility measures that we consider.

4.2 Robust RV Estimators

4.2.1 The “all” RV Estimator

The “all” estimator is equal to the standard realized volatility applied to the highest sampling frequency possible, i.e., the summation of the most finely sampled squared returns. The quadratic form representation is obviously,

$$RV_t^{all}(\underline{h}) \equiv RV_t(\underline{h}) = \sum_{i=1}^{1/\underline{h}} r_{t-1+i\underline{h}}^{(h)2} = V R_t(\underline{h})^\top Q^{all}(\underline{h}) V R_t(\underline{h}), \quad (4.8)$$

where

$$\begin{aligned} q_{ij}^{all}(\underline{h}) &= 1 \quad \text{for } i = j \\ &= 0 \quad \text{otherwise.} \end{aligned} \quad (4.9)$$

Of course, this estimator is not robust to market microstructure noise and, in fact, from the discussion above we would expect it to behave very poorly as an estimator for IV_t for small values of \underline{h} . However, the $RV_t^{all}(\underline{h})$ estimator serves an important role in the definition of some of the alternative estimators that we consider below.

4.2.2 The Sparse RV Estimator

The sparse realized volatility estimator similarly equals the traditional realized volatility estimator $RV_t(h)$, where h is assumed to be a multiple of \underline{h} ; i.e. $h = \underline{h}n_h$ where n_h is an integer. The quadratic representation for this estimator takes the form,

$$RV_t^{sparse}(h) = \sum_{i=1}^{1/h} r_{t-1+ih}^{(h)2} = VR_t(\underline{h})^\top Q^{sparse}(h) VR_t(\underline{h}), \quad (4.10)$$

where

$$\begin{aligned} q_{ij}^{sparse}(h) &= 1 \quad \text{for } i = j, \\ &= 1 \quad \text{for } i \neq j, (s-1)n_h + 1 \leq i, j \leq sn_h, s = 1, \dots, 1/h, \\ &= 0 \quad \text{otherwise.} \end{aligned} \quad (4.11)$$

As discussed in the previous section, by choosing h sufficiently large, this estimator can be rendered robust to the noise but it comes at the cost of decreasing the accuracy of the resulting estimate for IV_t . However, a new more desirable volatility estimator may be obtained by combining the $RV_t^{all}(\underline{h})$ and different $RV_t^{sparse}(h)$ estimators.

4.2.3 The Average RV Estimator

Following Zhang, Mykland and Ait-Sahalia (2005), we define the average realized volatility estimator as the mean of several sparse estimators. In particular, define the n_h distinct sparse estimators initiated respectively at $0, \underline{h}, 2\underline{h}, \dots, (n_h - 1)\underline{h}$, through the equation,

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

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

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

In terms of the quadratic form representation we have,

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

where

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

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

$$RV_t^{average}(h) = \frac{1}{n_h} \sum_{k=0}^{n_h-1} RV_t^{sparse}(h, k) = VR_t(\underline{h})^\top Q^{sparse}(h, k) VR_t(\underline{h}), \quad (4.15)$$

where

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

Whereas the standard sparse estimator only directly uses part of the sample, the average estimator more effectively exploits the data by extending the same estimator to each subgrid partition while retaining the associated robustness to noise for appropriately large values of h .

4.2.4 Two-Scale RV Estimator

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

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

The two-scale estimator may then be expressed as,

$$RV_t^{TS}(h) = RV_t^{average}(h) - \bar{n}\underline{h}RV_t^{all} = VR_t(\underline{h})^\top Q^{TS}(h) VR_t(\underline{h}), \quad (4.18)$$

where

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

In contrast to all of the previously defined estimators, the two-scale estimator is formally consistent for IV_t as $h \rightarrow 0$ under the noise assumptions discussed in Section 2.

4.2.5 The Adjusted Two-Scale RV Estimator

A simple correction for the number of terms entering into each of the sums defining the two-scale estimator suggests the following adjusted version,

$$RV^{TS-adj}(h) = (1 - \bar{n}\underline{h})^{-1} RV^{(TS)}(h) = VR_t(\underline{h})^\top Q^{TS-Adj}(h) VR_t(\underline{h}), \quad (4.20)$$

where

$$q_{ij}^{TS-adj}(h) = (1 - \bar{n}\underline{h})^{-1} q_{ij}^{TS}(h). \quad (4.21)$$

More elaborate multi-scale estimators have also been considered by Zhang (2006). However, the basic principle of these is similar to that of the two-scale estimators discussed above, and we do not analyze any of these further here.

4.2.6 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. Specifically,

$$\begin{aligned} RV_t^{Zhou}(h) &= \sum_{i=1}^{1/h} r_{t-1+ih}^{(h)2} + \sum_{i=2}^{1/h} r_{t-1+ih}^{(h)} r_{t-1+(i-1)h}^{(h)} + \sum_{i=1}^{1/h-1} r_{t-1+ih}^{(h)} r_{t-1+(i+1)h}^{(h)} \\ &= VR_t(\underline{h})^\top Q^{Zhou}(h) VR_t(\underline{h}), \end{aligned} \quad (4.22)$$

where

$$\begin{aligned} q_{ij}^{Zhou}(h) &= 1 \quad \text{for } i = j \\ &= 1 \quad \text{for } |i - j| = 1, (s-1)n_h + 1 \leq i, j \leq sn_h, s = 1, \dots, 1/h \\ &= 0 \quad \text{otherwise.} \end{aligned} \quad (4.23)$$

Note that on defining Zhou's estimator for the highest frequency \underline{h} , the expression for the components in $Q^{Zhou}(\underline{h})$ reduces to,

$$\begin{aligned} q_{ij}^{Zhou}(\underline{h}) &= 1 \quad \text{if } i = j \quad \text{or } |i - j| = 1, \\ &= 0 \quad \text{otherwise.} \end{aligned} \quad (4.24)$$

This is the version of the estimator that we use in the numerical results reported on below. This estimator has also previously been analyzed by Zumbach, Corsi and Trapletti (2002).

4.2.7 The Kernel-Based RV Estimator

The estimator discussed in the previous section may be seen as a special case of the general kernel type estimators developed by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006). Specifically, let $K(\cdot)$ and L denote the kernel and the bandwidth, respectively. The kernel-based realized volatility estimator is then defined by,

$$RV_t^{Kernel}(K(\cdot), L) = RV_t(\underline{h}) + \sum_{l=1}^L K\left(\frac{l-1}{L}\right) CRV_t(l, h), \quad (4.25)$$

where

$$CRV_t(l, h) = \sum_{i=1+l}^{1/h} r_{t-1+ih}^{(h)} r_{t-1+(i-l)h}^{(h)} + \sum_{i=1}^{1/h-l} r_{t-1+ih}^{(h)} r_{t-1+(i+l)h}^{(h)} \quad (4.26)$$

This estimator is readily expressed in quadratic form as,

$$RV_t^{Kernel}(K(\cdot), L) = VR_t(\underline{h})^\top Q^{Kernel}(K(\cdot), L) VR_t(\underline{h}), \quad (4.27)$$

where

$$\begin{aligned} q_{ij}^{Kernel}(K(\cdot), L) &= 1 \quad \text{for } i = j \\ &= K\left(\frac{l-1}{L}\right) \quad \text{for } |i - j| = l, \\ &= 0 \quad \text{otherwise.} \end{aligned} \quad (4.28)$$

In the specific calculations below, we use the modified Tukey-Hanning kernel advocated by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006),

$$K(x) = (1 - \cos\pi(1 - x)^2)/2. \quad (4.29)$$

However, the estimator that we actually use differs slightly from theirs, which in contrast to equation (4.26) adds returns outside the $[t - 1, t]$ time interval,

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

Since our analysis is explicitly focused on forecasting, we purposely do not want to include any returns beyond time t in the construction of the realized volatility for the $[t - 1, t]$ time-interval. Of course, this difference does not have any impact on the consistency and asymptotic distribution of the estimator. Also, we did not implement the optimal bandwidth selection procedure developed by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006), but instead simply fixed $L = h/\underline{h} - 1$, corresponding to the bandwidth implicitly used for the two-scale estimators. It is possible that alternative bandwidth choices might improve upon the performance measures reported below. Similarly, other kernels might work better in a forecasting context. We will not explore these issues further here, but our general analytical framework readily allows for such direct comparisons of different kernel-based realized volatility estimators across both kernel and bandwidth choices.

This completes our discussion of the robust realized volatility estimators which we will consider in the construction of the forecasts. However, before turning to a discussion of the forecast performance, we initially present, in the next section, a brief summary of the distributional characteristics of the different measures.

4.3 Distribution of Robust RV Measures

The analytical representation of the various cross-moments among the class of return variation estimators for models within the ESV class renders it feasible to directly compare their properties along various dimensions, even in the presence of noise. We continue to focus on Model 1 and 2 for a moderate as well as somewhat higher level of noise.

Table 4 provides the mean, variance and mean-squared-error for the various measures of integrated variance. The “all” estimator in principle entails using the about highest possible sampling frequency. We here settle on $h = 1/1440$, or 1-minute sampling in a 24-hour market, as a sensible choice for the shortest return interval. However, we also note that since the noise is calibrated to the findings in Hansen and Lunde (2006) for the equity market, the sampling may more appropriately be thought of as producing 15-second returns.

As predicted, the associated “all” estimator is badly inflated by the effect of microstructure noise. Even for moderate level of noise, the estimator on average attains a value almost four times as large as the underlying integrated variance. For the larger noise level, the bias produces an average estimates that is more than tenfold the true average return variation. In essence, as a direct estimator for the integrated variance, the measure is useless. Moving on to the sparse estimator based on $h = 1/288$, corresponding to 5-minute sampling in a 24-hour market, or 1-minute for a six hour equity trading day, we continue to see a significant upward bias in the associated realized volatility measure, although it has dropped sharply compared to the “all” estimator. A further reduction in sampling frequency would result in a less biased estimator and this may certainly be a sensible strategy in practice. For example, one could sample 78 times per day, corresponding to the relatively popular choice of 5-minute sampling for equity return series. Nonetheless, we retain the high underlying sampling frequency for the sparse estimator to explore more cleanly the implications of the noise-induced bias for predictive ability of these measures compared to the largely unbiased and robust measures discussed below. The last estimator constructed directly from the standard realized volatility measure is the average estimator appearing in the third row. The impact on the bias is negligible but, as expected, the averaging reduces the sampling variability associated with the discretization error, and in turn provides a small improvement in the overall mean-squared error compared to the regular sparse estimator.

Moving on to the estimators constructed with a view towards robustness to noise, we first consider the basic two-scale estimator. It shows an appreciable downward average bias in the associated realized variation estimates. The size of this bias is partially an artifact of the choice of sampling frequency for the sparse estimator relative to the “all” estimator, and would shrink if the sparse estimator is implemented for a lower sampling frequency. However, this issue is not an inherent problem for the approach as it is readily fixed through the simple multiplicative adjustment factor invoked by the adjusted two-scale estimator. In fact, the latter estimator is close to unbiased. Finally, the Zhou estimator, specifically designed for the type of noise analyzed here, and the kernel estimator are also practically unbiased. It is also interesting to note that all estimators, except for the (unadjusted) two-scale estimator, have about the same variance in the top panel, implying that the noise has little impact on the sampling variability of the estimators but instead predominantly affects the levels of the various measures differentially. Of course, this feature is not invariant to the level of the noise as the variances are much more dispersed in the lower panel, reflecting the noisier environment in effect there. In fact, the average estimator, the adjusted two-scale estimator and the kernel estimator are generally more stable, while the Zhou and sparse estimators appear to perform much worse in this regard.

Table 4 to be inserted here.

Table 5 explores how correlated the various estimates of return variation are. Since the actual integrated variance is also included in the table, this provides a first impression of the potential forecast performance of each measure, as a high correlation with the current volatility level, everything else equal, should translate into a good forecast. For the moderate noise level, we find the measures separating into two distinct groups. The “all”, sparse and Zhou estimators fail to match the performance of the remainder in terms of relatively close replication of the ideal integrated variance measure. It is noteworthy that the average estimator mimics the variation of the underlying return variation quite well in spite of its sizeable bias. On the other hand, the Zhou estimator may be close to unbiased but the increased sampling variability arising from adding two covariance terms to the sparse estimator to bias-correct may well prove costly in terms of predictive accuracy which is an issue we explore in the following sections.

For the considerably noisier environment with $\lambda = 0.5\%$ the discrepancies grow much larger between the two groups, indicating the importance of using an appropriate estimator that reflect the characteristics of the market setting. Moreover, within the poorer performing group, we now find the Zhou estimator looking decidedly less attractive than the sparse estimator. Another notable feature is that the kernel estimator starts performing less well than the alternatives among the better performing measures. Since the kernel estimator may be calibrated to closely replicate the two-scale estimator this simply reflects a less than optimal implementation of the kernel estimator within this noisier setting. It serves as a reminder that non-trivial performance gains may be achieved through careful choice of bandwidth (and kernel) for the various estimators. However, we do not pursue this issue any further given the relatively positive findings we are able to report in terms of general forecast performance in the following section.

Table 5 to be inserted here.

Our final summary comparison of the alternative return variation measures is provided in Table 6, which reports the population autocorrelations. Given the strong serial correlation of the underlying integrated variance series in Models 1 and 2, one should expect the estimators that more closely mimic this series to also exhibit a higher degree of persistence than the inherently noisier measures of return variation. This is exactly what the table shows. The ranking in terms of correlation with the integrated variance measure from Table 5 is preserved in terms of the relative strength of persistence reported in Table 6.

Table 6 to be inserted here.

4.4 True Forecasting Performance of Robust RV Measures

We now compare the actual forecast performance of the optimal linear forecasts constructed from the alternative return variation measures. In a direct extension of the previous findings for the regular realized volatility measures in Table 1, combining the results from Propositions 2.1 and 4.1, it is possible to analytically determine the true underlying population R^2 's. Given the wide variety of alternatives, we focus on only one version of each estimator. Also, we do not claim that the results represent the best possible implementation for any of the procedures, e.g., the sparse estimator may be improved by determining the optimal sampling frequency rather than simply relying on $h = 1/288$, for the more noisy setting the use of additional lags may provide further benefits, the two-scale estimator may well be improved through different sampling frequencies for the “all” and sparse estimators, as already previously noted, the kernel estimator can surely be improved through a more suitable bandwidth choice. Nonetheless, the findings illustrate the power of the ESV framework to summarize the performance of any specific forecasting design and compare it to relevant alternatives. Moreover, the results reported below are sufficiently impressive compared to the inherent predictability of the realized variation for each of the models, established in the top three rows of Table 1, that further experimentation to provide a marginal improvement in predictability will not alter the qualitative conclusions.

Table 7 to be inserted here.

Table 7 summarizes our results across horizons corresponding to an approximate daily, weekly and monthly period. Given the comparative properties of the estimators described in the previous section, the outcome is not particularly surprising. In all cases, the measures most highly correlated with the true underlying return variation also provide the best basis for forecasts. Hence, the forecasts generated by the average estimator are uniformly best, but they are barely different from the forecasts based on the two-scale estimators and, for the moderate noise scenario, the kernel estimator. Perhaps more striking is the relatively strong performance of the forecasts associated with the sparse estimator within the moderate noise setting. Likewise, the fall-off for any of the remaining forecasts is not dramatic under the realistic moderate noise setting. In fact, comparing the R^2 values with the degree of inherent predictability in these models and what may be achieved with feasible estimators under ideal noise-free conditions, as reported in Table 1, the performance of the entire range of alternative forecasts based on sensible realized variation measures is quite impressive. Of course, the performance for some of these forecasts deteriorates significantly when the noise level increases and there is an obvious gain to construct forecasts from appropriate integrated variance estimates in this situation. Nonetheless, the feasible gains for the sparse estimator, obtained by reducing the sampling frequency to around $h = 1/96$ and exploiting

lagged measures in the procedure, will again render forecasts based on the simpler estimator relatively competitive. Even so, the evidence suggests that the comparatively simple average estimator of daily return variation is an excellent starting point for volatility forecasting.

The superior performance of the average estimator may, of course, be an artifact of the *i.i.d.* noise assumption, although we still expect it to be a good candidate for the dependent noise case as well. The main difficulty in exploring this conjecture in a more systematic manner is that there is currently little consensus in the literature regarding the type of dependence in the noise process. It is also evident from prior research that the dependence structure is very different across intraday return series constructed from transaction prices versus bid-ask midpoint quotes, see, e.g., Hansen and Lunde (2006). On the other hand, once a tractable representation for the noise structure has been provided, the current framework offers a viable approach for exploring this scenario as well, as we briefly discuss below in Section 5.

4.5 Practically Feasible Forecasting Performance of Robust RV Measures

The integrated volatility appearing on the left-hand-side in the ideal Mincer-Zarnowitz regressions discussed in the previous section is, of course, latent. In practice the integrated volatility is typically replaced by some realized volatility measure, as in,

$$\overline{RM}_{t+1:t+m}(h) = a + bRM_t(h) + \eta_{t+m}, \quad (4.30)$$

where $\overline{RM}_t(h)$ and $RM_t(h)$ denote two possibly different realized measures, resulting in the regression R^2 ,

$$R^2 = \frac{(\text{Cov}[\overline{RM}_{t+1:t+m}(h), RM_t(h)])^2}{\text{Var}[\overline{RM}_{t+1:t+m}(h)]\text{Var}[RM_t(h)]}. \quad (4.31)$$

The following proposition provides closed form expressions for the requisite covariance term which, when combined with the previous expressions for the variances, may be used in numerically calculating the R^2 .

Proposition 4.2 *Let the discrete-time noise contaminated returns be determined by an ESV model and the relationship in equation (2.8). Let $\overline{RM}_t(h)$ and $RM_t(h)$ denote two realized volatility measures as defined in equation (4.1) with corresponding quadratic form weights q_{ij} and \bar{q}_{ij} , respectively. Then,*

$$\begin{aligned} \text{Cov}[\overline{RM}_{t+1}(h), RM_t(h)] = & \\ & \sum_{i=1}^{1/h} \sum_{k=1}^{1/h} \bar{q}_{ii} q_{kk} \left(\sum_{n=1}^p \frac{a_n^2}{\lambda_n^2} (1 - \exp(\lambda_n \underline{h}))^2 \exp(-\lambda_n(1 + (i - k - 1)\underline{h})) \right) \\ & + \bar{q}_{11} q_{h^{-1}\underline{h}^{-1}} (K_u - 1) V_u^2, \end{aligned} \quad (4.32)$$

and for $m > 1$,

$$\begin{aligned} \text{Cov}[\overline{RM}_{t+m}(h), RM_t(h)] = \\ \sum_{i=1}^{1/h} \sum_{k=1}^{1/h} \bar{q}_{ii} q_{kk} \left(\sum_{n=1}^p \frac{a_n^2}{\lambda_n^2} (1 - \exp(\lambda_n \underline{h}))^2 \exp(-\lambda_n(m + (i - k - 1)\underline{h})) \right). \end{aligned} \quad (4.33)$$

Proof: See Appendix II.

Exploiting the above results along with the previous techniques for analytic computation of the relevant R^2 values, we now provide the actual feasible performance measures that, ideally, may be obtained from the various forecast procedures discussed in the previous section. The findings are summarized in Table 8. We only report the relevant figures for the one-step-ahead forecasts due to the increased number of cross-comparisons that we provide.¹² As before, the relative rankings are preserved over the longer horizons. Table 8 conveys the now familiar picture. The average estimator dominates uniformly across all the feasible future realized return variation measures explored as right-hand-side variables in the predictive regressions. Moreover, the rankings from before are preserved everywhere across the alternative return generating processes (models) and noise levels. It is also evident that the use of more precise ex-post estimators for the integrated variance improves the measured degree of predictability and allows the regressions to convey in a reasonable sense the true underlying relationship as captured in Table 7. As a rather extreme example, consider Model 2 under the higher $\lambda = 0.5\%$ noise level. Using the average estimator as the basis for the forecast and as the ex-post proxy for future return variation realizations, one obtains an R^2 of 41% compared to the actual one-step-ahead R^2 of about 53% given in Table 7. In contrast, exploiting the Zhou estimator in both capacities instead results in an R^2 of less than 4%. Obviously, the specific figures are dependent on the model specification and noise structure assumed here, but it illustrates how the issue of observed versus underlying true predictability is crucially important in properly interpreting empirical studies in this area.¹³

Table 8 to be inserted here.

5 Extensions

All results generated so far are based on a number of simplifying assumption about the baseline diffusion in equation (2.1) and the structure of the noise in equation (2.8). In

¹²Also 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 $\text{Cov}[IV_{t+1}, RM_t(h)] \neq \text{Cov}[IV_{t+1}, IV_t]$ for $q_{ii} \neq 1$.

¹³As previously noted, ABM (2005) provide a technique for more formally converting the observed degree of predictability into an estimate of the higher true predictability through a fairly simple and practical procedure.

this section we point towards potential directions for extensions which relax these basic assumptions but retain the tractable analytical framework. Specifically, we provide, in turn, suggestions for ways to accommodate correlated noise processes, a leverage effect, and a drift term.

5.1 Correlated Noise

Our previous findings rest on the assumption that the noise in the observed price process induced by the market microstructure frictions result in an *i.i.d.* error term. However, the empirical evidence reported in Hansen and Lunde (2006) suggests that the noise may in fact be serially correlated over time. This recognition also motivates the estimator developed by Aït-Sahalia, Mykland and Zhang (2005b) which is designed to be consistent in the presence of serially correlated noise. Moreover, the kernel approach of Barndorff-Nielsen, Hansen, Lunde and Shephard (2006) also explicitly allows for dependent noise structures.

One approach to accommodate a more general noise structure is to assume that the error in the observed price process vis-a-vis the true latent price consists of the previously analyzed *i.i.d.* component as well as a stationary continuous-time serially correlated process of finite variation. Since the quadratic variation of the *i.i.d.* process is unbounded, the traditional realized volatility estimator will necessarily be inconsistent in this setting. However, the additional serially correlated components further complicate the analysis. In particular, suppose that the second component, \tilde{u}_t , is determined by the process,

$$\tilde{u}_t = \int_{-\infty}^t A(t-u)d\tilde{W}_u, \quad (5.1)$$

where \tilde{W}_u denotes a standard Brownian motion, assumed to be independent of the other stochastic processes in the system, i.e., the Brownian motions driving the price and volatility as well as the *i.i.d.* noise, and $A(\cdot)$ denotes a bounded real function in the sense that $\int_0^\infty A^2(u)du < \infty$.¹⁴ The modified version of the discrete-time return equation (2.8) then takes the form,

$$r_t^{(h)} = r_t^{*(h)} + e_t^{(h)} + \tilde{e}_t^{(h)}, \quad (5.2)$$

where

$$\tilde{e}_t^{(h)} \equiv \tilde{u}_t - \tilde{u}_{t-h}. \quad (5.3)$$

It is readily apparent that unless $A(\cdot)$ is identically equal to zero, the discretely observed returns will no longer adhere to a simple MA(1) correlation structure. However, following

¹⁴The Ornstein-Uhlenbeck (OU) process considered in Aït-Sahalia, Mykland and Zhang (2005a), $d\tilde{u}_t = -\tilde{\kappa}\tilde{u}_tdt + \tilde{\sigma}d\tilde{W}_t$, corresponds to the special case $A(s) = \tilde{\sigma}\exp(-\tilde{\kappa}s)$.

Feunou, Garcia, Meddahi, Tedongap (2006) the results in Proposition 2.1 may be extended to allow for non-trivial $A(\cdot)$ functions and general \tilde{u}_t error structures. In particular, one may show, for $i \geq j$,

$$\begin{aligned} Var[r_{t-1+ih}^{(h)}] &= a_0 h + 2V_u + \int_0^h A^2(u) du + \int_h^\infty [A(u) - A(u-h)]^2 du, \\ Cov[r_{t-1+ih}^{(h)}, r_{t-1+jh}^{(h)}] &= -V_u \delta_{i-j,1} \\ &+ \int_0^\infty (2A(u + (i-j)h) - A(u + (i-j+1)h) - A(u + (i-j-1)h)) A(u) du. \end{aligned}$$

We do not develop these ideas further here, but the formulas given above, along with related expressions for the higher order moments, immediately set the stage for investigating (analytically) the impact of quite general error structures for the different realized volatility measures and forecast regressions analyzed in the preceding sections.

5.2 Leverage Effect

Another simplifying assumption maintained in all of the analytical and numerical results discussed above concerns the absence of any leverage effect; i.e., the assumption that the Brownian motions driving the price and volatility processes are independent. Although this may be a viable assumption for some financial asset returns, notably foreign exchange, it is arguably not realistic for other series, e.g., equity index returns. Meanwhile, as shown in Meddahi (2002b) the ESV setting may be extended to allow for the derivation of the relevant moments of the non-contaminated returns, $r_{t-1+ih}^{*(h)}$, in the presence of a non-zero leverage effect. Merely to illustrate, we note that the unconditional fourth moment in this case may be expressed as,

$$E[(r_{t-1+ih}^{*(h)})^4] = 3a_0^2 h^2 + 6 \sum_{n=1}^p \frac{a_n^2}{\lambda_n^2} (\exp(-\lambda_n h) + \lambda_n h - 1) + 12\rho^2 C(h),$$

where the leverage coefficient, ρ , refers to the (instantaneous) correlation between the Brownian motions driving the price and volatility process, and the $C(h)$ function is explicitly defined in Meddahi (2002b). Again, this formula, along with related expressions for other conditional and unconditional moments of $r_{t-1+ih}^{*(h)}$, could be adapted in exploring the impact of leverage. However, we will not pursue these results any further here.

5.3 Drift

All of our previous results rule out any drift in the latent non-contaminated price process. This simplifying assumption is unlikely to materially affect any of the results related to the daily, weekly and monthly realized volatility measures and forecasts discussed above.

Nonetheless, the aforementioned extensions of the basic ESV setup developed in Meddahi (2000b) readily allows for a non-zero constant drift. For instance, including both leverage and a constant drift term, μdt , the fourth unconditional moment takes the form

$$E[(r_{t-1+ih}^{*(h)})^4] = 3a_0^2h^2 + 6 \sum_{n=1}^p \frac{a_n^2}{\lambda_n^2} (\exp(-\lambda_n h) + \lambda_n h - 1) + 12\rho^2 C(h) \\ + \mu^4 h^4 + 6\mu^2 a_0 h^2 + 12\mu\rho h \sum_{n=1}^p \frac{a_n e_{n,0}}{\lambda_n} (\exp(-\lambda_n h) + \lambda_n h - 1),$$

where $e_{n,0}$ is again explicitly defined in Meddahi (2002b).

Other generalizations of our basic framework are, of course, possible. However, we believe the most interesting extensions from a practical empirical perspective relate to the impact of correlated noise and leverage effects. The suggested representations along with the tools developed in the preceding sections provide a starting point for a direct analytical exploration and quantification of such effects in future work.

6 Conclusion

This paper extends existing analytic methods to the construction and assessment of volatility forecasts for continuous-time diffusion models to the empirically important case of market microstructure noise. The procedures work generally within the broad ESV class of models, which includes most of the popular volatility diffusion in current use, and may be adapted to accommodate many other empirically relevant features. We illustrate the techniques by applying them 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 varying degrees of noise with the theoretical upper bounds for the degree of predictability based on optimal (infeasible) forecasts. Under realistic scenarios, we find that it is feasible to produce quite precise forecasts but many aspects of the implementation of the forecasting schemes require careful evaluation of the underlying market structure and data availability in order to design the most effective procedures.

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

Appendix I: ESV Representations

This appendix provides a brief summary of the ESV representation for each of the three benchmark volatility models; more detailed discussions are available in Meddahi (2001) and ABM (2004).

Model M1 - GARCH Diffusion The GARCH diffusion in equation (2.22) is readily expressed as an ESV model by defining the state variable,

$$df_t = \kappa(\theta - f_t)dt + \sigma f_t dW_t^{(2)},$$

and the function $g(x) = x$. Assuming that the variance of σ_t^2 is finite, it follows that

$$\sigma_t^2 = a_0 + a_1 P_1(f_t),$$

where $a_0 = \theta$, $a_1 = \theta\sqrt{\psi/(1-\psi)}$, $\psi = \sigma^2/2\kappa$, and the first eigenfunction for f_t is affine,

$$P_1(x) = \frac{\sqrt{1-\psi}}{\theta\sqrt{\psi}}(x - \theta),$$

with corresponding eigenvalue $\lambda_1 = \kappa$.

Model M2 - Two-Factor Affine The two-factor affine model in (2.23) may be written in the form of an ESV model by defining the state variables,

$$df_{j,t} = \kappa_j(\alpha_j + 1 - f_{j,t})dt + \sqrt{2\kappa_j}\sqrt{f_{j,t}}dW_t^{(j+1)}, \quad j = 1, 2,$$

where $\alpha_j = (2\kappa_j\theta_j/\eta_j^2) - 1$, and the $f_{j,t}$'s are related to the $\sigma_{j,t}$'s by,

$$f_{j,t} = \frac{2\kappa_j}{\eta_j^2}\sigma_{j,t}^2, \quad j = 1, 2.$$

The eigenfunctions associated with $f_{j,t}$ are given by the Laguerre polynomials $L_n^{(\alpha_j)}(f_{j,t})$, $n = 0, 1, \dots$, with corresponding eigenvalues $\lambda_{j,n} = \kappa_j n$. Moreover,

$$\sigma_{j,t}^2 = \tilde{a}_{j,0} + \tilde{a}_{j,1}L_1^{(\alpha_j)}(f_{j,t})$$

where $\tilde{a}_{j,0} = \theta_j$ and $\tilde{a}_{j,1} = -\sqrt{\theta_j}\eta_j/\sqrt{2\kappa_j}$, so that,

$$\sigma_t^2 = a_{0,0} + a_{1,0}L_1^{(\alpha_1)}(f_{1,t}) + a_{0,1}L_1^{(\alpha_2)}(f_{2,t}),$$

with $a_{0,0} = \tilde{a}_{1,0} + \tilde{a}_{2,0}$, $a_{1,0} = \tilde{a}_{1,1}$ and $a_{0,1} = \tilde{a}_{2,1}$.

Model M3 - Log-Normal Diffusion The log-normal diffusion in (2.24) may be expressed in the form of an ESV model by defining the state variable,

$$df_t = -\kappa f_t dt + \sqrt{2\kappa} dW_t^{(2)},$$

where,

$$f_t = \frac{\sqrt{2\kappa}}{\sigma}(\log \sigma_t^2 - \theta).$$

The eigenfunctions associated with f_t are given by the Hermite polynomials $H_n(f_t)$, $n = 0, 1, \dots$, with corresponding eigenvalues $\lambda_n = \kappa n$, so that,

$$\sigma_t^2 = \sum_{n=0}^{\infty} a_n H_n(f_t),$$

where,

$$a_n = \exp\left(\theta + \frac{\sigma^2}{4\kappa}\right) \frac{(\sigma/\sqrt{2\kappa})^n}{\sqrt{n!}}.$$

Appendix II: Technical Proofs

Proof of Proposition 2.1. In the absence of any drift, $E[r_{t+ih}^{*(h)}] = 0$ and $Var[r_{t+ih}^{*(h)}] = a_0 h$ (see, e.g., Meddahi, 2002b). Now give 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

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

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

$$\begin{aligned} r_i r_j r_k r_l &= (r_i^* + e_i)(r_j^* + e_j)(r_k^* + e_k)(r_l^* + e_l) \\ &= r_i^* r_j^* r_k^* r_l^* + r_i^* r_j^* e_k r_l^* + r_i^* r_j^* r_k^* e_l + r_i^* r_j^* e_k e_l + r_i^* e_j r_k^* r_l^* + r_i^* e_j e_k r_l^* + r_i^* e_j r_k^* e_l + r_i^* e_j e_k e_l \\ &\quad + e_i r_j^* r_k^* r_l^* + e_i r_j^* e_k r_l^* + e_i r_j^* r_k^* e_l + e_i r_j^* e_k e_l + e_i e_j r_k^* r_l^* + e_i e_j e_k r_l^* + e_i e_j r_k^* e_l + e_i e_j e_k e_l. \end{aligned}$$

The returns are independent with 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,

$$\begin{aligned} E[r_i r_j r_k r_l] &= E[r_i^* r_j^* r_k^* r_l^*] + E[r_i^* r_j^*] E[e_k e_l] + E[r_i^* r_l^*] E[e_j e_k] + E[r_i^* r_k^*] E[e_j e_l] \\ &\quad + E[e_i e_k] E[r_j^* r_l^*] + E[e_i e_l] E[r_j^* r_k^*] + E[e_i e_j] E[r_k^* r_l^*] + E[e_i e_j e_k e_l]. \end{aligned} \quad (\text{A.1})$$

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),

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

Equations (3.7) and (3.10) in Meddahi (2002b) now imply

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

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

$$\begin{aligned} e_i e_j e_k e_l &= (u_i - u_{i-1})(u_j - u_{j-1})(u_k - u_{k-1})(u_l - u_{l-1}) \\ &= u_i u_j u_k u_l - u_i u_j u_k u_{l-1} - u_i u_j u_{k-1} u_l + u_i u_j u_{k-1} u_{l-1} \\ &\quad - u_i u_{j-1} u_k u_l + u_i u_{j-1} u_k u_{l-1} + u_i u_{j-1} u_{k-1} u_l - u_i u_{j-1} u_{k-1} u_{l-1} \\ &\quad - u_{i-1} u_j u_k u_l + u_{i-1} u_j u_k u_{l-1} + u_{i-1} u_j u_{k-1} u_l - u_{i-1} u_j u_{k-1} u_{l-1} \\ &\quad + u_{i-1} u_{j-1} u_k u_l - u_{i-1} u_{j-1} u_k u_{l-1} - u_{i-1} u_{j-1} u_{k-1} u_l + u_{i-1} u_{j-1} u_{k-1} u_{l-1}. \end{aligned}$$

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

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

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

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

Therefore,

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

Now combining (A.2), (A.3), and (A.4) results (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

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

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 $h > 0$,*

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

Proof of Lemma A.1. We have

$$\text{Cov} \left[\int_a^b \sigma_u^2 du, \int_c^d \sigma_u^2 du \right] = \sum_{1 \leq n, m \leq p} a_n a_m E \left[\int_a^b P_n(f_u) du \int_c^d P_m(f_u) du \right] \\ = \sum_{1 \leq n, m \leq p} a_n a_m E \left[\int_a^b P_n(f_u) du \int_c^d E[P_m(f_u) \mid f_\tau, \tau \leq b] du \right] \\ = \sum_{1 \leq n, m \leq p} a_n a_m \int_a^b E[P_n(f_u) P_m(f_b)] du \int_c^d \exp(-\lambda_m (u - b)) du \\ = \sum_{1 \leq n, m \leq p} a_n a_m \int_a^b \exp(-\lambda_n (b - u)) \delta_{n,m} du \frac{[1 - \exp(-\lambda_m (d - c))]}{\lambda_m} \exp(-\lambda_m (c - b)) \\ = \sum_{1 \leq n, m \leq p} a_n a_m \delta_{n,m} \frac{[1 - \exp(-\lambda_n (b - a))]}{\lambda_n} \frac{[1 - \exp(-\lambda_m (d - c))]}{\lambda_m} \exp(-\lambda_n (c - b)),$$

i.e., (A.5).

In order to prove (2.20) note that the independence of the noise with the volatility process implies that

$$\begin{aligned} E[IV_t r_{t-1+ih}^{*(h)} r_{t-1+jh}^{*(h)}] &= E[IV_t r_{t-1+ih}^{*(h)} r_{t-1+jh}^{*(h)}] + E[IV_t] E[e_{t-1+ih}^{(h)} e_{t-1+jh}^{(h)}] \\ &= E[IV_t r_{t-1+ih}^{*(h)} r_{t-1+jh}^{*(h)}] + a_0(2V_u \delta_{i,j} - V_u \delta_{|i-j|,1}). \end{aligned} \quad (\text{A.6})$$

We will now prove

$$\begin{aligned} E[IV_t r_{t-1+ih}^{*(h)} r_{t-1+jh}^{*(h)}] &= \sum_{n=1}^p \frac{a_n^2}{\lambda_n^2} [2 - \exp(-\lambda_n(i-1)h) - \exp(-\lambda_n(1-ih))] [1 - \exp(-\lambda_n h)] \\ &\quad + a_0^2 h + 2 \sum_{n=1}^p \frac{a_n^2}{\lambda_n^2} [\exp(-\lambda_n h) + \lambda_n h - 1] \quad \text{if } i = j, \\ &= 0 \quad \text{otherwise,} \end{aligned} \quad (\text{A.7})$$

which, combined with (A.6), leads to (2.20).

Given the path of the volatility, the returns are independent. Hence,

$$E[IV_t r_{t-1+ih}^{*(h)} r_{t-1+jh}^{*(h)}] = E[IV_t E[r_{t-1+ih}^{*(h)} r_{t-1+jh}^{*(h)} \mid \sigma_\tau, \tau \leq t]] = \delta_{i,j} E \left[IV_t \int_{t-1+(i-1)h}^{t-1+ih} \sigma_u^2 du \right]. \quad (\text{A.8})$$

On the other hand, Lemma A.1 implies

$$\begin{aligned} E \left[IV_t \int_{t-1+(i-1)h}^{t-1+ih} \sigma_u^2 du \right] &= E \left[\left(\int_{t-1}^{t-1+(i-1)h} \sigma_u^2 du + \int_{t-1+(i-1)h}^{t-1+ih} \sigma_u^2 du + \int_{t-1+ih}^t \sigma_u^2 du \right) \int_{t-1+(i-1)h}^{t-1+ih} \sigma_u^2 du \right] \\ &= \sum_{n=1}^p \frac{a_n^2}{\lambda_n^2} [2 - \exp(-\lambda_n(i-1)h) - \exp(-\lambda_n(1-ih))] [1 - \exp(-\lambda_n h)] + a_0^2 h(1-h) \\ &\quad + E \left[\left(\int_{t-1+(i-1)h}^{t-1+ih} \sigma_u^2 du \right)^2 \right]. \end{aligned}$$

Equations (12) and (15) in ABM (2004) imply

$$E \left[\left(\int_{t-1+(i-1)h}^{t-1+ih} \sigma_u^2 du \right)^2 \right] = a_0^2 h^2 + 2 \sum_{n=1}^p \frac{a_n^2}{\lambda_n^2} [\exp(-\lambda_n h) + \lambda_n h - 1].$$

Consequently,

$$\begin{aligned} E \left[IV_t \int_{t-1+(i-1)h}^{t-1+ih} \sigma_u^2 du \right] &= \sum_{n=1}^p \frac{a_n^2}{\lambda_n^2} [2 - \exp(-\lambda_n(i-1)h) - \exp(-\lambda_n(1-ih))] [1 - \exp(-\lambda_n h)] \\ &\quad + a_0^2 h + 2 \sum_{n=1}^p \frac{a_n^2}{\lambda_n^2} [\exp(-\lambda_n h) + \lambda_n h - 1]. \end{aligned} \quad (\text{A.9})$$

Combining (A.8) and (A.9) results in (A.7), which completes the proof of (2.20).

Similar arguments used in the proof of (2.20) lead to

$$\begin{aligned} E[IV_{t+1:t+m} r_{t-1+ih}^{*(h)} r_{t-1+jh}^{*(h)}] &= E[IV_{t+1:t+m} r_{t-1+ih}^{*(h)} r_{t-1+jh}^{*(h)}] + E[IV_{t+1:t+m}] E[e_{t-1+ih}^{(h)} e_{t-1+jh}^{(h)}] \\ &= \delta_{i,j} E \left[IV_{t+1:t+m} \int_{t-1+(i-1)h}^{t-1+ih} \sigma_u^2 du \right] + a_0 m (2V_u \delta_{i,j} - V_u \delta_{|i-j|,1}). \end{aligned} \quad (\text{A.10})$$

By using Lemma A.1 it follows that

$$E \left[IV_{t+1:t+m} \int_{t-1+(i-1)h}^{t-1+ih} \sigma_u^2 du \right] = a_0^2 h m + \sum_{i=1}^p \frac{a_i^2}{\lambda_i^2} [1 - \exp(-\lambda_n h)] [1 - \exp(-\lambda_n m)] \exp(-\lambda_n(1-ih)),$$

(A.11)

which, combined with (A.10), lead to (2.21). ■

Proof of Proposition 3.1. To begin, it follows readily from (2.8) that

$$RV_t(h) = RV_t^*(h) + \sum_{i=1}^{1/h} e_{t-1+ih}^{(h)2} + 2 \sum_{i=1}^{1/h} r_{t-1+ih}^{*(h)} e_{t-1+ih}^{(h)}. \quad (\text{A.12})$$

Equation (A.12) and the independence of the noise with the volatility process implies that

$$Cov[IV_{t+1:t+m}, RV_{t-l}(h)] = Cov[IV_{t+1:t+m}, RV_{t-l}^*(h)],$$

i.e., the first equation in (3.2). The second equation in (3.2) is a direct consequence equation (32) in ABM (2004).

To prove (3.3), note that the independence assumption for the noise u_t and the $\{\log(S_t), \sigma_t\}$ processes, along with the zero mean return, imply that

$$Var[RV_t(h)] = Var[RV_t^*(h)] + Var \left[\sum_{i=1}^{1/h} e_{t-1+ih}^{(h)2} \right] + 4Var \left[\sum_{i=1}^{1/h} r_{t-1+ih}^{*(h)} e_{t-1+ih}^{(h)} \right]. \quad (\text{A.13})$$

Similarly,

$$Var \left[\sum_{i=1}^{1/h} e_{t-1+ih}^{(h)2} \right] = \frac{1}{h} Var[e_{t-1+ih}^{(h)2}] + 2 \sum_{1 \leq i < j \leq 1/h} Cov[e_{t-1+ih}^{(h)2}, e_{t-1+jh}^{(h)2}].$$

On the other hand,

$$Var[e_{t-1+ih}^{(h)2}] = Var[e_h^{(h)2}] = 2(K_u + 1)V_u^2$$

while

$$Cov[e_{t-1+ih}^{(h)2}, e_{t-1+jh}^{(h)2}] = Cov[e_{ih}^{(h)2}, e_{jh}^{(h)2}] = \delta_{i,(j-1)}(K_u - 1)V_u^2.$$

Therefore,

$$Var \left[\sum_{i=1}^{1/h} e_{t-1+ih}^{(h)2} \right] = \frac{1}{h} 2(K_u + 1)V_u^2 + 2 \left(\frac{1}{h} - 1 \right) (K_u - 1)V_u^2 = 2V_u^2 \left(\frac{2K_u}{h} - K_u + 1 \right). \quad (\text{A.14})$$

The noise process $e_t^{(h)}$ and the returns $\{r_{t-1+ih}^{*(h)}\}$ are independent and the returns are uncorrelated. Hence,

$$Var \left[\sum_{i=1}^{1/h} r_{t-1+ih}^{*(h)} e_{t-1+ih}^{(h)} \right] = \frac{1}{h} Var \left[r_{t-1+ih}^{*(h)} e_{t-1+ih}^{(h)} \right] = \frac{1}{h} E[r_{t-1+ih}^{*(h)2}] E[e_{t-1+ih}^{(h)2}] = 2E[\sigma_t^2] V_u, \quad (\text{A.15})$$

given that $E[r_{t-1+ih}^{*(h)2}] = hE[\sigma_t^2]$. Combining (A.13), (A.14), and (A.15), achieves (3.3).

We will now prove (3.4) and (3.5). The same arguments used in the proof of (A.13) imply that

$$\begin{aligned} \forall m \geq 1, Cov[RV_{t+m}(h), RV_t(h)] &= Cov[RV_{t+m}^*(h), RV_t^*(h)] + Cov \left[\sum_{i=1}^{1/h} e_{t+m-1+ih}^{(h)2}, \sum_{i=1}^{1/h} e_{t-1+ih}^{(h)2} \right] \\ &\quad + 4Cov \left[\sum_{i=1}^{1/h} r_{t+m-1+ih}^{*(h)} e_{t+m-1+ih}^{(h)}, \sum_{i=1}^{1/h} r_{t-1+ih}^{*(h)} e_{t-1+ih}^{(h)} \right]. \end{aligned} \quad (\text{A.16})$$

But, it follows also that

$$\begin{aligned} \forall i, j, Cov \left[r_{t+m-1+ih}^{*(h)} e_{t+m-1+ih}^{(h)}, r_{t-1+jh}^{*(h)} e_{t-1+jh}^{(h)} \right] &= E \left[r_{t+m-1+ih}^{*(h)} e_{t+m-1+ih}^{(h)} r_{t-1+jh}^{*(h)} e_{t-1+jh}^{(h)} \right] \\ &= E \left[r_{t+m-1+ih}^{*(h)} r_{t-1+jh}^{*(h)} \right] E \left[e_{t+m-1+ih}^{(h)} e_{t-1+jh}^{(h)} \right] \\ &= 0. \end{aligned}$$

Therefore,

$$Cov \left[\sum_{i=1}^{1/h} r_{t+m-1+ih}^{*(h)} e_{t+m-1+ih}^{(h)}, \sum_{i=1}^{1/h} r_{t-1+ih}^{*(h)} e_{t-1+ih}^{(h)} \right] = 0. \quad (\text{A.17})$$

We also have

$$Cov \left[\sum_{i=1}^{1/h} e_{t+m-1+ih}^{(h)2}, \sum_{i=1}^{1/h} e_{t-1+ih}^{(h)2} \right] = \delta_{m,1} Cov \left[e_{t+h}^{(h)2}, e_t^{(h)2} \right] = \delta_{m,1} Cov \left[e_t^2, e_t^2 \right] = \delta_{m,1} (K_u - 1) V_u^2. \quad (\text{A.18})$$

Combining (A.16), (A.17), and (A.18), results in

$$\forall m \geq 1, Cov[RV_{t+m}(h), RV_t(h)] = Cov[RV_{t+m}^*(h), RV_t^*(h)] + \delta_{m,1} (K_u - 1) V_u^2,$$

i.e., (3.4) when $m = 1$ and (3.5) when $m > 1$. This achieves the proof of Proposition 3.1. ■

Proof of Proposition 4.2. We have

$$Cov[\overline{RM}_{t+m}(h), RM_t(h)] = \sum_{1 \leq i, j, k, l \leq 1/h} \bar{q}_{ij} q_{kl} Cov[r_{t+m+ih}^{(h)} r_{t+m+jh}^{(h)}, r_{t+kh}^{(h)} r_{t+lh}^{(h)}]. \quad (\text{A.19})$$

When $m > 1$, (A.19) combined with the first part of (2.19) leads to (4.33). When $m = 1$, (A.19) combined with the second and third parts of (2.19) leads to (4.32). This achieves the proof of Proposition 4.2. ■

Table 1: R^2 for Integrated Variance Forecasts

λ	1/h	Model	M1			M2			M3		
			Horizon	1	5	20	1	5	20	1	5
0%	1440	$R^2(Best)$	0.977	0.891	0.645	0.830	0.586	0.338	0.989	0.945	0.807
		$R^2(IV_t)$	0.955	0.871	0.630	0.689	0.445	0.214	0.977	0.934	0.796
		$R^2(IV_t, 4)$	0.957	0.874	0.632	0.698	0.446	0.227	0.979	0.936	0.797
	288	$R^2(RV_t^*(h))$	0.950	0.867	0.627	0.679	0.439	0.211	0.974	0.931	0.793
		$R^2(RV_t^*(h), 4)$	0.951	0.868	0.627	0.685	0.450	0.224	0.974	0.931	0.793
	96	$R^2(RV_t^*(h))$	0.932	0.851	0.615	0.641	0.414	0.199	0.960	0.918	0.781
		$R^2(RV_t^*(h), 4)$	0.934	0.852	0.616	0.642	0.429	0.216	0.963	0.920	0.784
	48	$R^2(RV_t^*(h))$	0.891	0.813	0.588	0.563	0.364	0.175	0.927	0.886	0.754
		$R^2(RV_t^*(h), 4)$	0.908	0.829	0.599	0.580	0.395	0.202	0.946	0.904	0.770
	1	$R^2(RV_t^*(h))$	0.836	0.762	0.551	0.476	0.307	0.148	0.881	0.843	0.717
		$R^2(RV_t^*(h), 4)$	0.883	0.805	0.582	0.519	0.360	0.186	0.929	0.889	0.757
		$R^2(RV_t^*(h), 19)$	0.122	0.111	0.081	0.031	0.020	0.010	0.157	0.150	0.128
0.1%	1440	$R^2(RV_t(h))$	0.360	0.329	0.238	0.072	0.054	0.029	0.452	0.432	0.369
		$R^2(RV_t(h), 4)$	0.493	0.450	0.325	0.092	0.074	0.043	0.639	0.611	0.523
		$R^2(RV_t(h))$	0.896	0.817	0.591	0.547	0.353	0.170	0.936	0.895	0.762
	288	$R^2(RV_t(h), 4)$	0.911	0.831	0.601	0.569	0.388	0.199	0.950	0.908	0.773
		$R^2(RV_t(h))$	0.908	0.828	0.599	0.581	0.375	0.181	0.943	0.901	0.767
	96	$R^2(RV_t(h), 4)$	0.917	0.837	0.605	0.594	0.402	0.205	0.953	0.911	0.776
		$R^2(RV_t(h))$	0.873	0.797	0.576	0.525	0.339	0.163	0.914	0.874	0.744
	48	$R^2(RV_t(h), 4)$	0.899	0.821	0.593	0.553	0.379	0.195	0.941	0.900	0.766
		$R^2(RV_t(h))$	0.821	0.749	0.541	0.450	0.291	0.140	0.870	0.832	0.708
	1	$R^2(RV_t(h), 4)$	0.877	0.800	0.578	0.501	0.349	0.182	0.926	0.885	0.754
		$R^2(RV_t(h))$	0.122	0.111	0.081	0.031	0.020	0.010	0.157	0.150	0.127
		$R^2(RV_t(h), 19)$	0.360	0.328	0.237	0.071	0.053	0.029	0.452	0.432	0.368
0.5%	1440	$R^2(RV_t(h))$	0.492	0.449	0.325	0.091	0.074	0.042	0.638	0.611	0.522
		$R^2(RV_t(h))$	0.446	0.407	0.294	0.123	0.080	0.038	0.554	0.529	0.451
		$R^2(RV_t(h), 4)$	0.711	0.649	0.469	0.222	0.164	0.088	0.811	0.776	0.661
	288	$R^2(RV_t(h))$	0.719	0.656	0.474	0.300	0.194	0.093	0.800	0.765	0.651
		$R^2(RV_t(h), 4)$	0.837	0.764	0.552	0.395	0.283	0.149	0.903	0.864	0.736
	96	$R^2(RV_t(h))$	0.772	0.704	0.509	0.365	0.236	0.113	0.839	0.802	0.683
		$R^2(RV_t(h), 4)$	0.858	0.782	0.566	0.443	0.313	0.165	0.915	0.875	0.746
	48	$R^2(RV_t(h))$	0.858	0.782	0.566	0.443	0.313	0.165	0.915	0.875	0.746
		$R^2(RV_t(h), 4)$	0.750	0.684	0.495	0.349	0.225	0.108	0.817	0.781	0.665
	1	$R^2(RV_t(h))$	0.849	0.775	0.560	0.431	0.306	0.161	0.909	0.869	0.740
		$R^2(RV_t(h))$	0.121	0.110	0.080	0.031	0.020	0.010	0.155	0.149	0.126
		$R^2(RV_t(h), 4)$	0.357	0.326	0.236	0.070	0.052	0.028	0.449	0.430	0.366
1.0%	1440	$R^2(RV_t(h), 19)$	0.491	0.448	0.324	0.090	0.073	0.042	0.637	0.610	0.521
		$R^2(RV_t(h))$	0.178	0.163	0.118	0.037	0.024	0.012	0.249	0.238	0.203
		$R^2(RV_t(h), 4)$	0.458	0.418	0.302	0.084	0.063	0.034	0.586	0.560	0.478
	288	$R^2(RV_t(h))$	0.466	0.425	0.308	0.133	0.086	0.041	0.574	0.548	0.467
		$R^2(RV_t(h), 4)$	0.723	0.660	0.477	0.234	0.172	0.092	0.820	0.784	0.669
	96	$R^2(RV_t(h))$	0.620	0.566	0.409	0.222	0.143	0.069	0.714	0.683	0.581
		$R^2(RV_t(h), 4)$	0.798	0.728	0.526	0.329	0.238	0.127	0.875	0.837	0.713
	48	$R^2(RV_t(h))$	0.648	0.591	0.428	0.248	0.160	0.077	0.735	0.702	0.598
		$R^2(RV_t(h), 4)$	0.810	0.739	0.534	0.352	0.254	0.135	0.882	0.844	0.719
	1	$R^2(RV_t(h))$	0.119	0.109	0.079	0.030	0.019	0.009	0.154	0.147	0.125
		$R^2(RV_t(h), 4)$	0.354	0.323	0.234	0.069	0.052	0.028	0.446	0.427	0.364
		$R^2(RV_t(h), 19)$	0.488	0.445	0.322	0.089	0.072	0.041	0.635	0.608	0.519

Table 2: R^2 for 'Optimally' Sampled Intraday Returns

λ	$1/h_1$	$1/h_2$	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
1.0%	15.3	48.7	$R^2(RV_t(h_1))$	0.567	0.517	0.374
			$R^2(RV_t(h_1), 4)$	0.775	0.707	0.511
			$R^2(RV_t(h_2))$	0.648	0.591	0.428
			$R^2(RV_t(h_2), 4)$	0.810	0.739	0.534
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
1.0%	14.1	43.1	$R^2(RV_t(h_1))$	0.199	0.128	0.062
			$R^2(RV_t(h_1), 4)$	0.307	0.223	0.119
			$R^2(RV_t(h_2))$	0.249	0.161	0.077
			$R^2(RV_t(h_2), 4)$	0.353	0.255	0.135
Model M3						
0.1%	74.0	520	$R^2(RV_t(h_1))$	0.901	0.861	0.733
			$R^2(RV_t(h_1), 4)$	0.936	0.895	0.762
			$R^2(RV_t(h_2))$	0.946	0.905	0.770
			$R^2(RV_t(h_2), 4)$	0.955	0.913	0.777
0.5%	25.3	104	$R^2(RV_t(h_1))$	0.762	0.728	0.620
			$R^2(RV_t(h_1), 4)$	0.891	0.852	0.726
			$R^2(RV_t(h_2))$	0.839	0.802	0.683
			$R^2(RV_t(h_2), 4)$	0.916	0.875	0.746
1.0%	16.0	52.0	$R^2(RV_t(h_1))$	0.657	0.628	0.535
			$R^2(RV_t(h_1), 4)$	0.855	0.817	0.696
			$R^2(RV_t(h_2))$	0.735	0.703	0.598
			$R^2(RV_t(h_2), 4)$	0.882	0.844	0.719

Table 3: R^2 for Optimally Combined Intraday Squared Returns

		Model	M1			M2		
λ	1/h	Horizon	1	5	20	1	5	20
0.1%	1440	$R^2(RV_t(h))$	0.896	0.817	0.591	0.547	0.353	0.170
		$R^2(\text{Optimal})$	0.897	0.819	0.592	0.562	0.359	0.171
	288	$R^2(RV_t(h))$	0.908	0.828	0.599	0.581	0.376	0.180
		$R^2(\text{Optimal})$	0.910	0.830	0.600	0.600	0.382	0.182
	96	$R^2(RV_t(h))$	0.873	0.797	0.576	0.525	0.339	0.163
		$R^2(\text{Optimal})$	0.874	0.798	0.577	0.537	0.343	0.164
	48	$R^2(RV_t(h))$	0.821	0.749	0.541	0.450	0.290	0.140
		$R^2(\text{Optimal})$	0.821	0.749	0.542	0.457	0.293	0.140
	1	$R^2(RV_t(h))$	0.122	0.111	0.080	0.031	0.020	0.010
		$R^2(\text{Optimal})$	0.122	0.111	0.080	0.031	0.020	0.010
0.5%	1440	$R^2(RV_t(h))$	0.446	0.407	0.294	0.123	0.080	0.038
		$R^2(\text{Optimal})$	0.446	0.407	0.294	0.124	0.080	0.038
	288	$R^2(RV_t(h))$	0.719	0.656	0.474	0.300	0.194	0.093
		$R^2(\text{Optimal})$	0.719	0.656	0.475	0.303	0.195	0.093
	96	$R^2(RV_t(h))$	0.772	0.704	0.509	0.365	0.235	0.113
		$R^2(\text{Optimal})$	0.772	0.705	0.510	0.369	0.237	0.114
	48	$R^2(RV_t(h))$	0.750	0.684	0.495	0.349	0.225	0.108
		$R^2(\text{Optimal})$	0.750	0.684	0.495	0.353	0.227	0.109
	1	$R^2(RV_t(h))$	0.121	0.110	0.080	0.030	0.020	0.009
		$R^2(\text{Optimal})$	0.121	0.110	0.080	0.030	0.020	0.009

Table 4: Mean, Variance and MSE of RV Measures

Model	M1			M2		
	Mean	Variance	MSE	Mean	Variance	MSE
IV_t	0.636	0.168	0.168	0.504	0.0263	0.0263
$\lambda = 0.1\%$						
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}	0.507	0.110	0.127	0.402	0.018	0.029
RV_t^{TS-Adj}	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
$\lambda = 0.5\%$						
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}	0.507	0.117	0.133	0.402	0.023	0.033
RV_t^{TS-Adj}	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

Table 5: Correlations of RV Measures

	IV_t	RV_t^{all}	RV_t^{sparse}	$RV_t^{average}$	RV_t^{TS}	RV_t^{Zhou}	RV_t^{Kernel}
Model M1							
$\lambda = 0.1\%$							
IV_t	1.00	0.969	0.975	0.989	0.986	0.971	0.986
RV_t^{all}	-	1.00	0.958	0.972	0.956	0.932	0.954
RV_t^{sparse}	-	-	1	0.986	0.984	0.962	0.981
$RV_t^{average}$	-	-	-	1.00	0.998	0.976	0.995
RV_t^{TS}	-	-	-	-	1.00	0.978	0.996
RV_t^{Zhou}	-	-	-	-	-	1.00	0.983
RV_t^{Kernel}	-	-	-	-	-	-	1.00
$\lambda = 0.5\%$							
IV_t	1.00	0.684	0.868	0.964	0.958	0.745	0.932
RV_t^{all}	-	1.00	0.690	0.766	0.601	0.335	0.551
RV_t^{sparse}	-	-	1.00	0.901	0.878	0.667	0.838
$RV_t^{average}$	-	-	-	1.00	0.974	0.740	0.930
RV_t^{TS}	-	-	-	-	1.00	0.802	0.963
RV_t^{Zhou}	-	-	-	-	-	1.00	0.816
RV_t^{Kernel}	-	-	-	-	-	-	1.00
Model M2							
$\lambda = 0.1\%$							
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
$\lambda = 0.5\%$							
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

Table 6: Autocorrelations of RV Measures

Lag	RV_t^{all}	RV_t^{sparse}	$RV_t^{average}$	RV_t^{TS}	RV_t^{Zhou}	RV_t^{Kernel}
Model M1						
$\lambda = 0.1\%$						
1	0.917	0.929	0.956	0.949	0.921	0.949
2	0.885	0.897	0.923	0.917	0.889	0.917
5	0.797	0.808	0.831	0.825	0.800	0.825
10	0.669	0.678	0.698	0.693	0.672	0.693
20	0.472	0.478	0.492	0.488	0.473	0.488
$\lambda = 0.5\%$						
1	0.457	0.736	0.907	0.896	0.542	0.849
2	0.441	0.711	0.876	0.865	0.523	0.819
5	0.397	0.640	0.789	0.779	0.471	0.737
10	0.333	0.537	0.662	0.654	0.395	0.619
20	0.235	0.378	0.466	0.461	0.279	0.436
Model M2						
$\lambda = 0.1\%$						
1	0.659	0.700	0.773	0.756	0.673	0.521
2	0.516	0.548	0.606	0.592	0.527	0.408
5	0.320	0.340	0.376	0.367	0.327	0.253
10	0.203	0.215	0.238	0.232	0.207	0.160
20	0.094	0.100	0.111	0.108	0.096	0.089
$\lambda = 0.5\%$						
1	0.149	0.361	0.640	0.618	0.197	0.521
2	0.116	0.283	0.502	0.484	0.154	0.407
5	0.072	0.176	0.311	0.300	0.096	0.253
10	0.046	0.111	0.197	0.190	0.061	0.160
20	0.021	0.052	0.092	0.088	0.028	0.088

Table 7: R^2 for Integrated Variance Forecasts

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

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						
$\lambda = 0.1\%$						
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
$\lambda = 0.5\%$						
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						
$\lambda = 0.1\%$						
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
$RV^{av.}$	0.095	0.231	0.410	0.396	0.126	0.333
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
$\lambda = 0.5\%$						
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

References

- Aït-Sahalia, Y., L.P. Hansen and J. Scheinkman (2006), "Operator Methods for Continuous-Time Markov Models," in Y. Aït-Sahalia and L.P. Hansen (Eds.), *Handbook of Financial Econometrics*, forthcoming.
- Aït-Sahalia and L. Mancini (2006), "Out of Sample Forecasts of Quadratic Variation," unpublished manuscript, Princeton University.
- Aït-Sahalia, P.A. Mykland and L. Zhang (2005a), "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise," *Review of Financial Studies*, 18, 351-416.
- Aït-Sahalia, P.A. Mykland and L. Zhang (2005b), "Ultra High Frequency Volatility Estimation with Dependent Microstructure Noise," unpublished manuscript, Princeton University.
- Andersen, T.G. and T. Bollerslev (1998), "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, 39, 885-905.
- Andersen, T.G., T. Bollerslev and F.X. Diebold (2006), "Parametric and Nonparametric Measurements of Volatility," in Y. Aït-Sahalia and L.P. Hansen (Eds.), *Handbook of Financial Econometrics*, forthcoming.
- Andersen, T.G., T. Bollerslev, P.F. Christoffersen and F.X. Diebold (2006), "Volatility Forecasting," in G. Elliott, C.W.J. Granger and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, North-Holland.
- Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2000), "Great Realizations," *Risk*, 13, 105-108.
- Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2001), "The Distribution of Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42-55.
- Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2003), "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579-625.
- Andersen, T.G., T. Bollerslev and S. Lange (1999), "Forecasting Financial Market Volatility: Sample Frequency vis-à-vis Forecast Horizon," *Journal of Empirical Finance*, 6, 457-477.
- Andersen, T.G., T. Bollerslev and N. Meddahi (2004), "Analytic Evaluation of Volatility Forecasts," *International Economic Review*, 45, 1079-1110.

- Andersen, T.G., T. Bollerslev and N. Meddahi (2005), "Correcting the Errors: Volatility Forecast Evaluation Using High-Frequency Data and Realized Volatilities," *Econometrica*, 73, 279-296.
- Areal, N.M.P.C. and S.J. Taylor (2002), "The Realized Volatility of FTSE-100 Futures Prices," *Journal of Futures Market*, 22, 627-648.
- Bandi, F. and J. Russell (2005a), "Separating Microstructure Noise from Volatility," *Journal of Financial Economics*, 79, 655-692.
- Bandi, F. and J. Russell (2005b), "Microstructure Noise, Realized Volatility, and Optimal Sampling," unpublished manuscript, University of Chicago.
- Bandi, F., J. Russell and Y. Zhu (2006), "Using High-Frequency Data in Dynamic Portfolio Choice," *Econometric Reviews*, forthcoming.
- Barndorff-Nielsen, O.E. and N. Shephard (2001), "Non-Gaussian OU based Models and some of their uses in Financial Economics," with discussion, *Journal of the Royal Statistical Society, B*, 63, 167-241.
- Barndorff-Nielsen, O.E. and N. Shephard (2002a), "Econometric Analysis of Realised Volatility and its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society, B*, 64, 253-280.
- Barndorff-Nielsen, O.E. and N. Shephard (2002b), "Estimating Quadratic Variation Using Realised Variance," *Journal of Applied Econometrics*, 5, 457-477.
- Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde and N. Shephard (2006), "Designing Realised Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise," unpublished manuscript, Oxford University.
- Chen, X., L.P. Hansen and J. Scheinkman (2005), "Principal Components and the Long Run," unpublished manuscript, University of Chicago.
- Corradi, V., W. Distaso and N.R. Swanson (2005), "Predictive Density Estimators for Daily Volatility Based on the Use of Realized Measures," unpublished manuscript, University of London.
- Corradi, V., W. Distaso and N.R. Swanson (2006), "Predictive Inference for Integrated Volatility," unpublished manuscript, University of London.
- Corsi, F. (2003), "A Simple Long Memory Model of Realized Volatility," unpublished manuscript, University of Southern Switzerland.
- Deo, R., C. Hurvich and Y. Lu (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. and G.G.J. Lee (1999), "A Permanent and Transitory Component Model for Stock Return Volatility," in R.F. Engle and H. White (Eds.), *Cointegration, Causality and Forecasting: a festschrift in Honour of Clive W.J. Granger*, Oxford University Press.
- Feunou, B., R. Garcia, N. Meddahi, and R. Tedongap (2006), "Estimation of Continuous Time Models based on Realized Measures: a Comparison of Methods." Work in progress, Université de Montréal.
- Garcia, R., M.A. Lewis, S. Pastorello and E. Renault (2001), "Estimation of Objective and Risk-Neutral Distributions Based on Moments of Integrated Volatility," unpublished manuscript, University of Montréal.
- Garcia, R. and N. Meddahi (2006), "Comment on Realized Variance and Market Microstructure Noise," *Journal of Business and Economic Statistics*, 24, 184-191.
- Ghysels, E., P. Santa-Clara and R. Valkanov (2006), "Predicting Volatility: Getting the Most out of Return Data Sampled at Different Frequencies," *Journal of Econometrics*, 131, 59-96.
- Ghysels, E. and A. Sinko (2006), "Volatility Forecasting and Microstructure Noise," unpublished manuscript, University of North Carolina.
- Hansen, P.R. and A. Lunde (2006), "Realized Variance and Market Microstructure Noise," *Journal of Business and Economic Statistics*, 24, 127-161.
- Hansen, L.P. and J. Scheinkman (1995), "Back to the Future: Generating Moment Implications for Continuous Time Markov Processes," *Econometrica*, 63, 767-804.
- Hull, J. and A. White (1987), "The Pricing of Options on Assets with Stochastic Volatilities," *Journal of Finance*, Vol XLII, 281-300.
- Koopman, S.J., B. Jungbacker and E. Hol (2005), "Forecasting Daily Variability of the S&P100 Stock Index using Historical, Realized and Implied Volatility Measures," *Journal of Empirical Finance*, 12, 445-475.
- Martens, M. (2002), "Measuring and Forecasting S&P500 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. (2006), "Volatility Forecast Comparison using Imperfect Volatility Proxies," London School of Economics working paper.
- Pong, S., M.B. Shackleton, S.J. Taylor and X. Xu (2004), "Forecasting Currency Volatility: A Comparison of Implied Volatilities and AR(FI)MA Models," *Journal of Banking and Finance*, 28, 2541-2563.
- Thomakos, D.D. and T. Wang (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*, forthcoming.
- Zhang, L., P.A. Mykland and Y. Aït-Sahalia (2005), "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," *Journal of the American Statistical Association*, 100, 1394-1411.
- Zhou, B. (1996), "High-Frequency Data and Volatility in Foreign Exchange Rates," *Journal of Business and Economic Statistics*, 14, 45-52.
- Zumbach, G., F. Corsi and A. Trapletti (2002), "Efficient Estimation of Volatility using High-Frequency Data," unpublished manuscript, Olsen & Associates, Zürich, Switzerland.