Journal of Econometrics 160 (2011) 257–271

Contents lists available at ScienceDirect

Journal of Econometrics

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

Volatility forecasting and microstructure noise[☆]

Eric Ghysels[a,b,*], Arthur Sinko[c,1]

[a] Department of Finance, Kenan-Flagler, School of Business, United States

[b] Department of Economics, University of North Carolina, Gardner Hall CB 3305, Gardner Hall CB 3305, Chapel Hill, NC 27599-3305, United States

[c] Economics, School of Social Sciences, Arthur Lewis Building University of Manchester, Manchester M13 9PL, United Kingdom

ARTICLE INFO

Article history:
Available online 6 March 2010

ABSTRACT

It is common practice to use the sum of frequently sampled squared returns to estimate volatility, yielding the so-called realized volatility. Unfortunately, returns are contaminated by market microstructure noise. Several noise-corrected realized volatility measures have been proposed. We assess to what extent correction for microstructure noise improves forecasting future volatility using a MIxed DAta Sampling (MIDAS) regression framework. We study the population prediction properties of various realized volatility measures, assuming *i.i.d*. microstructure noise. Next we study optimal sampling issues theoretically, when the objective is forecasting and microstructure noise contaminates realized volatility. We distinguish between conditional and unconditional optimal sampling schemes, and find that conditional optimal sampling seems to work reasonably well in practice.

1. Introduction

We study a regression prediction problem with volatility measures that are contaminated by microstructure noise and examine optimal sampling for the purpose of volatility prediction. The analysis is framed in the context of MIDAS regressions with regressors affected by microstructure noise. We investigate two topics: (1) the theoretical analysis and empirical performance of various volatility measures sampled at different frequencies, and (2) the performance of volatility measures corrected for independent noise compared to those that are corrected for dependent market microstructure noise.

Discussions about the impact of microstructure have mostly focused so far on *measurement* and therefore mean squared error and bias of various adjustments. In this paper, instead the focus is on prediction in a regression format, and therefore we can include estimators that are suboptimal in mean square error sense, since covariation with the predictor is what matters. Previously, the optimal sampling frequency was studied in terms of *MSE of estimators* in an asymptotic setting (Zhang et al., 2005) and for finite samples (Bandi and Russell, 2008b). In this respect, the issues discussed here differ from the existing literature.

We also conduct an extensive empirical study of forecasting with microstructure noise. We use the same data as in Hansen and Lunde (2006), namely the thirty Dow Jones Industrial Average (DJIA), from January 3, 2000 to December 31, 2004. The purpose of our empirical analysis is twofold. First, we verify whether the predictions from the theory hold in actual data samples. We find that is indeed the case. Second, we also implement optimal sampling schemes empirically and check the relevance of the theoretical derivations using real data. We distinguish between "conditional" and "unconditional" optimal sampling schemes, as in Bandi and Russell (2006). We find that "conditional" optimal sampling seems to work reasonably well in practice.

The topic of this paper has been studied by many authors independently and simultaneously. Garcia and Meddahi (2006) and Ghysels and Sinko (2006) discussed forecasting volatility and microstructure noise. Ghysels et al. (2007) provided further empirical evidence expanding on Ghysels and Sinko (2006). Aït-Sahalia and Mancini (2008) consider a number of stochastic volatility and jump–diffusions, including the Heston and log-volatility models, and study the relative performance of the two-scales realized (henceforth TSRV) estimator versus RV estimators. They provide simulation evidence showing that TSRV largely outperforms RV. They also report an empirical application which confirms their simulation results. We derive theoretical results for RV, TSRV, average over subsamples and Zhou (1996) estimators and study theoretically optimal sampling as well. For the most part we consider *i.i.d.* noise in our theoretical derivations. In addition, we also cover extensively empirical evidence on the topic.

☆ We thank Yacine Aït-Sahalia, Federico Bandi, Peter Hansen, Jeffrey Russell, and Per Mykland for their helpful comments, and Peter Hansen and Asger Lunde for providing their data. We also like to thank Neil Shephard for his detailed comments on an earlier draft of our paper as well as the Referees. All remaining errors are ours.

* Corresponding author at: Department of Economics, University of North Carolina, Gardner Hall CB 3305, Gardner Hall CB 3305, Chapel Hill, NC 27599-3305, United States. Tel.: +1 919 966 5325.

E-mail addresses: eghysels@unc.edu (E. Ghysels), arthur.sinko@manchester.ac.uk (A. Sinko).

1 Tel.: +44 161 275 4842.

0304-4076/$ – see front matter © 2010 Published by Elsevier B.V.
doi:10.1016/j.jeconom.2010.03.035

Andersen et al. (2011) considered independently several issues covered in this paper.

The paper is structured in the following way. Section 2 provides the theoretical underpinnings for our analysis. Section 3 describes the univariate MIDAS prediction models, the data and the empirical implementation of optimal sampling. Section 4 discusses the results. Section 5 concludes.

2. Volatility prediction and microstructure noise

We want to compare the forecasting performance of linear regression models with various realized volatility measures as regressors. To do so we need to study the population second-order moments of the these volatility measures with future realizations, i.e. the regressands in our analysis.

2.1. Description of realized measures

We use p to denote an observable, high-frequency log-price process, p^* to refer to the unobservable efficient log-price process, and η to denote the microstructure noise component that has mean 0 and variance σ_η^2. Also, we assume that the noise component and the efficient price component are independent. The relation between them is given by the equation

$$p_t = p_t^* + \eta_t. \tag{2.1}$$

We will assume that

$$\mathrm{d}p_t^* = \sigma_t \mathrm{d}W_t, \tag{2.2}$$

where W_t is a standard Brownian motion and the spot volatility process σ_t is predictable and has a continuous sample path. We first focus on equally-spaced data sampling (calendar data sampling). We denote by $M+1$ the number of prices associated with the finest equidistant grid $\mathcal{M}$ (say — every second) per period (say a day). Because of microstructure noise, we will also consider estimators that use a relatively "sparse grid" $\mathcal{M}_j^m$ and constructed using $\overline{M} = M/m$ returns.

We consider six realized volatility estimators. Their technical details are reported in Appendix A.1. The first estimator, the daily RV_j^m, defined in Eq. (A.2), is consistent in the absence of microstructure noise, but biased and inconsistent when noise is present. The second estimator, RV_{AC_1}, defined in Eq. (A.3), was studied by Zhou (1996) and Hansen and Lunde (2006). Under the assumption of *i.i.d.* microstructure noise it is unbiased, but inconsistent. The third estimator, RV_{TS} (Eq. (A.4)), is the two-scales estimator of Zhang et al. (2005). It consists of two parts. The first part is the average of a "fast-scale" RV_j^m measures. Since each measure uses only $\overline{M}$ intraperiod returns, potential improvement could be made by averaging over m different realized volatility estimators. The second part is a properly adjusted "slow-scale" realized volatility. It is introduced to compensate for the microstructure noise bias. Although the bias correction can significantly change estimated realized volatility, we will show in Sections 2.3 and 4 that it may not significantly change the mean square error of prediction. Nevertheless, we expect RV_{TS} to perform well as subsampling reduces the measurement error. This prompts another question, namely how good is a averaging estimator without the noise-correction? We write it as $\overline{\mathrm{RV}}^m$ and it is define in Eq. (A.5). The last two estimators we consider capture the fact that, in reality, microstructure noise can be serially correlated. The first estimator, RV_{TS_d} (Eq. (A.7)), is a modification of (A.4) and is proposed by Aït-Sahalia et al. (2005). They only differ in the definition of the bias-correction term. The last estimator, proposed by Hansen and Lunde (2006) and $\mathrm{RV}^{1\,\mathrm{tick}}_{ACNW_w}$ (Eq. (A.8)), is Bartlett kernel-based defined in tick-time instead of calendar-time.

2.2. Data generating processes

To derive population covariances of realized variance estimators we need to specify a data generating process for spot volatility. We assume that the variance is a function of a state variable which is a linear combination of the eigenfunctions of the infinitesimal generator associated with the state variable in continuous time. As noted by Renault (2009), the eigenfunction approach dates back to Granger and Newbold (1976) and was applied to volatility by Taylor (1982) and further developed by Meddahi (2001). Special cases of this setting include the log-normal and the square-root processes where the eigenfunctions are the Hermite and Laguerre polynomials, respectively. The eigenfunction approach has several advantages including the fact that any square integrable function may be written as a linear combination of the eigenfunctions and the implied dynamics of the variance and squared return processes have an ARMA representation. Therefore, one can easily compute forecasting formulas. This approach has been successfully used for such a purpose in a number of recent papers including, in the context of forecasting with microstructure noise, in independent work by Andersen et al. (2011).

We describe first some properties of the return process and microstructure noise. The observed log-price process is described by Eq. (2.1). For microstructure noise process η_t we rely on an *i.i.d.* assumption.[2] In particular:

Assumption 2.1. The η_t process is *i.i.d.* with $\mathrm{E}(\eta_t) = 0$, $\mathrm{Var}(\eta_t) = \sigma_\eta^2$, $\mathrm{E}\left(\eta_t^3\right) = 0$, $\mathrm{E}\left(\eta_t^4\right)/\mathrm{E}\left(\eta_t^2\right)^2 = \kappa$, η_t is independent from p_t^*. We assume that σ_t^2 is a continuous square-integrable function of a Markovian time-reversible stationary process f_t, namely: $\sigma_t^2 = a_0 + \sum_{i=1}^k a_i P_i(f_t)$ where a_i is real; $k \leq \infty$; $a_i = \mathrm{E}\left(\sigma^2(f_t)P_i(f_t)\right)$; $\forall k > 0$: $\mathrm{E}\left(P_i(f_{t+k})|f_t\right) = \mathrm{e}^{-\lambda_i k}P_i(f_t)$; $\mathrm{E}\left(P_i(f_t)\right) = \delta_{i,1}$; $\mathrm{E}\left(P_i(f_t)P_j(f_t)\right) = \delta_{i,j}$. Also, we assume that constituents of p^* (Eq. (2.2)) W and σ are independent processes (no leverage assumption).

Given this setting and Eqs. (2.1) and (A.1), observed returns are defined as $r_{t,h} = p_t^* - p_{t-h}^* + \eta_t - \eta_{t-h} = r_{t,h}^* + e_{t,h}$ and therefore in the absence of a drift $r_{t,h}^* = \int_{t-h}^t \sigma_t \mathrm{d}W_t$. For this setting we prove[3]:

Theorem 2.1. *Let Assumption 2.1 hold, then for the estimators* RV, $\mathrm{RV}_{AC_1}^m$, $\overline{\mathrm{RV}}$, *and* RV_{TS}*:*

$$\begin{aligned}
&\mathrm{Bias}\left(\mathrm{RV}_j^m\right) = \mathrm{Bias}\left(\overline{\mathrm{RV}}^m\right) = (1 - \overline{M}h)a_0 + 2\overline{M}\sigma_\eta^2;\\
&\mathrm{Bias}\left(\mathrm{RV}_{TS}\right) = \left(1 - \overline{M}h - \frac{1}{M}\right)a_0
\end{aligned} \tag{2.3}$$

$$\begin{aligned}
\mathrm{Var}(\mathrm{RV}_j^m) &= 2\sum_{i=1}^p \frac{a_i^2}{\lambda_i^2}\left(\mathrm{e}^{-\lambda_i \overline{M}h} - 1 + \lambda_i \overline{M}h\right) + 2\overline{M}a_0^2h^2\\
&\quad + 4\overline{M}\sum_{i=1}^p \frac{a_i^2}{\lambda_i^2}\left(\mathrm{e}^{-\lambda_i h} - 1 + \lambda_i h\right) + 8\overline{M}a_0 h\sigma_\eta^2\\
&\quad + (4\overline{M} - 2)(\kappa - 1)\sigma_\eta^4 + 4\overline{M}\sigma_\eta^4
\end{aligned} \tag{2.4}$$

$$\begin{aligned}
\mathrm{Var}(\mathrm{RV}_{AC_1}^m) &= (2.4) + 4\frac{\overline{M}^2}{(\overline{M}-1)}\left[(\kappa + 2)\sigma_\eta^4 + 4a_0h\sigma_\eta^2 + a_0^2h^2\right.\\
&\quad \left. + \sum_{i=1}^p \frac{a_i^2}{\lambda_i^2}\left(1 - \mathrm{e}^{-\lambda_i h}\right)^2\right] + 8\frac{\overline{M}^2(\overline{M}-2)}{(\overline{M}-1)^2}\sigma_\eta^4
\end{aligned}$$

2 Note that this assumption is relevant only for a relatively low-frequency setup. For discussion of microstructure noise properties see, for example, Hansen and Lunde (2006).

3 It should be noted that our results differ slightly from the findings the independently written paper by Andersen et al. (2011), as we are using only intraperiod prices to compute realized volatility estimators. As a result two noisy returns that span consecutive *periods* have independent noise components.

$$-8\overline{M}\left[(\kappa+1)\sigma_\eta^4+2a_0h\sigma_\eta^2\right] \tag{2.5}$$

$$\begin{aligned}
\mathrm{Var}\left(\overline{\mathrm{RV}}^m\right) &= \frac{8a_0h\overline{M}\sigma_\eta^2}{m}+\frac{(4\overline{M}-2)(\kappa-1)\sigma_\eta^4+4\overline{M}\sigma_\eta^4}{m} \\
&+\frac{\overline{M}}{m}\left(2h^2a_0^2+4\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(\mathrm{e}^{-\lambda_k h}-1+\lambda_k h\right)\right) \\
&\times\frac{1}{m^2}\left(\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\sum_{i=0}^{m-1}(1+2i)\left[\mathrm{e}^{-\lambda_k(1-2i/M)}-1\right.\right. \\
&\left.\left.+\lambda_k(1-2i/M)\right]\right)+\frac{4}{m^2}\sum_{i=1}^{m-1}\sum_{j=0}^{i-1} \\
&\times\left(\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left[1-\mathrm{e}^{-\lambda_k(1-2i/M)}\right]\left[1-\mathrm{e}^{-\lambda_k\left(\frac{i-j}{M}\right)}\right]\right) \\
&+\frac{2}{m^2}\sum_{i=1}^{m-1}\sum_{j=0}^{i-1}4(\overline{M}-1)\left\{\frac{(i-j)^2a_0^2}{2M^2}\right. \\
&\left.+\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(\mathrm{e}^{-\lambda_k\frac{i-j}{M}}-1+\lambda_k\frac{i-j}{M}\right)\right\} \\
&+\frac{2}{m^2}\sum_{i=1}^{m-1}\sum_{j=0}^{i-1}4\overline{M}\left\{\frac{(m-i+j)^2a_0^2}{2M^2}\right. \\
&\left.+\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(\mathrm{e}^{-\frac{\lambda_k(m-i+j)}{M}}-1+\lambda_k\frac{m-i-j}{M}\right)\right\}
\end{aligned} \tag{2.6}$$

$$\begin{aligned}
\mathrm{Var}\left(\mathrm{RV}_{TS}\right) &= (2.6)+\frac{\overline{M}^2}{M^2}\left(2\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(\mathrm{e}^{-\lambda_k}-1+\lambda_k\right)\right. \\
&\left.+\frac{2a_0^2}{M}+4M\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(\mathrm{e}^{-\lambda_k/M}-1+\lambda_k/M\right)\right) \\
&+\frac{\overline{M}^2}{M^2}\left(8a_0\sigma_\eta^2+(4M-2)(\kappa-1)\sigma_\eta^4+4M\sigma_\eta^4\right) \\
&-\frac{2\overline{M}}{M}\left(4\overline{M}\sigma_\eta^2/M+\left(4\overline{M}-2/m\right)(\kappa-1)\sigma_\eta^4\right) \\
&-\frac{2\overline{M}}{Mm}\sum_{j=0}^{m-1}\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left((2-\mathrm{e}^{-\lambda_k(m-1-j)/M}-\mathrm{e}^{-\lambda_k j/M})\right. \\
&\left.\times(1-\mathrm{e}^{-\lambda_k\overline{M}h})+(\mathrm{e}^{-\lambda_k\overline{M}h}-1+\lambda_k\overline{M}h)\right)-\frac{2\overline{M}(M-m)}{M} \\
&\times\left(2a_0^2/M^2+4\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}(\mathrm{e}^{-\lambda_k/M}-1+\lambda_k/M)\right).
\end{aligned} \tag{2.7}$$

Proof. See Appendix A.2. □

2.3. Population prediction properties and optimal sampling

We now turn to the computations of multiple correlation coefficients, or R^2, for single regressor equations projecting future integrated (realized) volatility onto various past RV measurements and a constant. By "approximately optimal[4]" we mean the following: to unify our analysis, we assume that $\mathrm{Cov}\left(\widehat{\mathrm{RV}}_{t+nh,nh},\widehat{\mathrm{RV}}_{t,h}\right)=\mathrm{Cov}\left(IV_{t+nh,nh},IV_{t,h}\right)$ for all estimators we consider in this section. In reality, however, this is only an approximation. Moreover, we (1) do not take into account overnight returns, and (2), neglect a bias of the averaging over subsamples estimators, that disappears only asymptotically.[5] In addition, we derive approximately optimal sampling frequencies in terms of minimization of *MSE of the prediction*, which is equivalent in this case to the optimal frequency maximization of the population R^2.

This analysis parallels that of Andersen et al. (2004), in particular, they show that for n-period ahead forecasts,

$$\begin{aligned}
R^2(\widehat{\mathrm{RV}}_{t+nh,nh},\widehat{\mathrm{RV}}_{t,h}) &= \mathrm{Cov}\left(\widehat{\mathrm{RV}}_{t+nh,nh},\widehat{\mathrm{RV}}_{t,h}\right)^2 \\
&\times\left[\mathrm{Var}\left(\widehat{\mathrm{RV}}_{t+nh,nh}\right)\mathrm{Var}\left(\widehat{\mathrm{RV}}_{t,h}\right)\right]^{-1}.
\end{aligned} \tag{2.8}$$

From the above equation we note that the relative R^2-performance of different $\mathrm{RV}_{t,h}$ measures only depends on the variances of $\widehat{\mathrm{RV}}_{t,h}$, since $\mathrm{Cov}\left(\widehat{\mathrm{RV}}_{t+nh,nh},\widehat{\mathrm{RV}}_{t,h}\right)=\mathrm{Cov}\left(IV_{t+nh,nh},IV_{t,h}\right)$ and $\mathrm{Var}\left(\widehat{\mathrm{RV}}_{t+nh,nh}\right)$ is fixed. As a result, whenever $\mathrm{Var}\left(\widehat{\mathrm{RV}}^A_{t,h}\right)\geq\mathrm{Var}\left(\widehat{\mathrm{RV}}^B_{t,h}\right)$ then $R^2(\widehat{\mathrm{RV}}_{t+nh,nh},\widehat{\mathrm{RV}}^A_{t,h})\leq R^2(\widehat{\mathrm{RV}}_{t+nh,nh},\widehat{\mathrm{RV}}^B_{t,h})$. The above setting also applies to multiple regressor cases, denoted by $R^2\left(\widehat{\mathrm{RV}}_{t+nh,nh};\widehat{\mathrm{RV}}_{t,h},\widehat{\mathrm{RV}}_{t-h,h},\ldots,\widehat{\mathrm{RV}}_{t-lh,h}\right)$, which will be relevant for the empirical analysis reported in the next sections. Defining the $l+1\times1$ vector $C(y_\tau,z_t,l)=(\mathrm{Cov}\,(y_\tau,z_t)\,,\mathrm{Cov}\,(y_\tau,z_{t-1})\,,\ldots,\mathrm{Cov}\,(y_\tau,z_{t-l}))'$ and the $(l+1)\times(l+1)$ matrix $M(z_t,l)$ with (i,j)th component $M(z_t,l)[i,j]=\mathrm{Cov}\left(z_t,z_{t+i-j}\right)$, we can then express the R^2 for a regression of $\mathrm{RV}_{t+nh,nh}$ onto a constant and the set of regressors $(\mathrm{RV}_{t,h},\mathrm{RV}_{t-h,h},\ldots,\mathrm{RV}_{t-lh,h})$, $l\geq0$ as:

$$\begin{aligned}
R^2\left(\mathrm{RV}_{t+nh,nh},\mathrm{RV}_{t,h},l\right) &= C\,(.)'\left[M\left(\mathrm{RV}_{t,h},l\right)\right]^{-1} \\
&\times C\,(.)\,/\mathrm{Var}\left(\mathrm{RV}_{t+nh,nh}\right)
\end{aligned} \tag{2.9}$$

where $C(.)=C\left(\mathrm{RV}_{t+nh,nh},\mathrm{RV}_{t,h},l\right)$.

Theorem 2.2. *For multiple regressions and arbitrary realized volatility estimators X, Y, and Z yielding R^2's in (2.9), and whenever $\mathrm{Cov}\left(\widehat{\mathrm{RV}}^i_{t+a,a},\widehat{\mathrm{RV}}^j_{t-\delta,b}\right)=\mathrm{Cov}\left(IV_{t+a,a},IV_{t-\delta,b}\right)$, $\forall\delta\geq0$ and $i,j=\{X,Y,Z\}$ we have that:*

$$\begin{aligned}
R^2\left(\widehat{\mathrm{RV}}^Z_{t+nh,nh},\widehat{\mathrm{RV}}^X_{t,h},l\right)\geq R^2\left(\widehat{\mathrm{RV}}^Z_{t+nh,nh},\widehat{\mathrm{RV}}^Y_{t,h},l\right) &\Leftrightarrow \mathrm{Var}\left(\widehat{\mathrm{RV}}^X_{t,h}\right) \\
&\leq \mathrm{Var}\left(\widehat{\mathrm{RV}}^Y_{t,h}\right).
\end{aligned}$$

Proof. See Appendix A.2. □

With the above result we are able to proceed with the comparison of the impact of various volatility measures on forecasting performance. To apply Theorem 2.2 to our framework, we have to assume that the covariances between daily estimators are equal to the covariances between daily integrated volatilities.

Consider two groups of estimators: group $A=\{\mathrm{RV},\mathrm{RV}_{AC_1}\}$ and group $B=\{\overline{\mathrm{RV}},\mathrm{RV}_{TS}\}$. We expect that group B estimators should have smaller variance both in noise-free and noisy environments. In a noise-free environment the variances are smaller given that the variance of the discretization noise (Eq. (A.21)) averaged over subsamples is smaller than the variance of $\overline{M}\mathrm{Var}\left(Z_{j/M+h,h}\right)$ (see Eq. (A.19)).

When microstructure noise is present, as the sampling frequency becomes large, the variances of group A estimators diverge

4 For convenience we will henceforth use the term "optimality" when "optimality in term of MSE of prediction" is discussed. In all other cases we will state explicitly when we mean MSE of estimation.

5 The exact covariances (assuming daily integrated volatility can be consistently estimated without overnight returns) are shown in Ghysels and Sinko (2009). All numerical results are obtained using these formulas.

to infinity. As a result, the R^2s of the regressions converge to 0. Group *B* estimators show a different pattern. These two estimators have asymptotically zero variance, and the two-scales correction only removes the bias from the estimator averaging over subsamples. However, if the bias is constant over time, it does not affect the predictive power of the regression as it only changes the intercept. Thus, the performance of group *B* estimators should be approximately the same, with the averaging over subsamples estimator performing better as the "sparse" grid $\overline{M}$ approaches M. Finally, in real-data applications, the variance of market microstructure noise is time-varying, which makes the two-scales estimator preferable. In particular, we have to emphasize the following. First, our analysis is valid as long as a regressand is unbiased, note though that in the simplest case we consider, microstructure noise bias does not change the prediction MSE and relative performance of different regressors. Note also, however, that such analysis cannot be applied directly to autoregressive models. For the autoregressive models there is a non-trivial bias-variance tradeoff. Second, our analysis is about picking the best regressor *given* some regressand. Hence, we do not answer the question which is the best regressand. Finally, we operate under the assumption that the noise is *i.i.d.*

The microstructure *i.i.d.* assumption impacts our analysis in two ways. First, in our derivations of variances we assume that market microstructure noise has zero autocorrelation. It is however well-documented in the literature that this does not hold for ultra-high frequencies.[6] This violation of our theoretical assumption creates a wedge between our results and the empirics. Second, in our regression analysis we assume that the variance and fourth moment of microstructure noise do not vary over time. During market crashes or severe liquidity shocks, however, the variance of microstructure noise tends to increase (see, for example, Aït-Sahalia and Yu (2009)). Violation of this creates a time-varying regressor bias. It should decrease the regression explanatory power of $\overline{\mathrm{RV}}$ relative to the RV_{TS} estimator and make predictions biased. However, our empirical analysis will show that there is no significant difference, for all frequencies we consider, in the explanatory power of $\overline{\mathrm{RV}}$ and RV_{TS} estimators. Thus, the *i.i.d.* assumption allows us to develop a simple and concise theory, and check qualitatively for what frequencies this theory works and whether a time-varying bias should be taken into account. To focus exclusively on a relative performance of different realized volatility measures, we use as a regressand the infeasible integrated variance.

To appraise the performance of the realized volatility estimators, we compare the population predictive power of the estimators using three models M1–M3 described in Andersen et al. (2004). Model M1 is a GARCH-diffusion model, M2 is a two-factor affine model, and M3 is a log-normal diffusion one (for details and parameters see Appendix A.2, Eqs. (A.24)–(A.26)).

Table 1 presents the sample results for models M1–M3. The results are reported for different frequencies, number of lags, zero and non-zero microstructure noise. Market microstructure noise is assumed to have *i.i.d.* with $\kappa = 3$, and two values of $\sigma_\eta^2 : \sigma_\eta^2 = 0.001$ (relevant for the frequently traded stocks we consider in the empirical part of the paper), and $\sigma_\eta^2 = 0.03$. The latter σ_η^2 is rather unrealistic. However, in reality, not only the variance of microstructure noise, but also the variance of fundamental returns varies across stocks, while in our analysis it is fixed for a given model. Aït-Sahalia and Yu (2009) show that noise-to-signal ratio (NSR) for traded stocks varies from virtually 0 (no noise) to 1 (no signal). Thus, the high value of σ_η^2 describes stocks with high NSR price process.

Each model, for each set of microstructure noise properties, has three sampling frequencies: five minutes, one minute and twenty seconds. These frequencies correspond accordingly to the low-frequency scales for RV_{TS} and $\overline{\mathrm{RV}}$, and sampling frequencies for the RV and Zhou estimators. The high-frequency scale for RV_{TS} and $\overline{\mathrm{RV}}$ is set to be one second. These sampling frequencies are selected to check the suitability of the forecasting procedure method and are commonly used in empirical studies utilizing the 5 min sampling frequency. The full version of the tables is available online in Ghysels and Sinko (2009). We consider three values for lags: 1, 15, 50; seven values of the variance of market microstructure noise $\sigma^2 = 0, 0.001, 0.005, 0.01, 0.015, 0.02, 0.025, 0.03$, and six values for the microstructure noise kurtosis $\kappa = 1.5, 2, 2.5, 3, 3.5, 4$. The main findings are the following:

In the noise-free environment the averaging over subsampling estimator performs the best across all estimators. The two-scales estimator produces slightly worse results. These two estimators outperform the RV and RV_{AC_1} estimators. Moreover, the RV estimator performs better than RV_{AC_1}. Infeasible linear regressions with integrated variance on the right hand side also perform better compared to the feasible ones. This finding goes along the same line as Theorem 2.2. Lower variance of the estimator implies higher R^2 assuming the all other assumptions hold. As the sampling frequency increases, the variance of the discretization noise decreases and R^2 of all estimators converge to the R^2 of the infeasible regression on integrated volatility (R^2_{IV}). For all the models, fifteen lags are sufficient and further increase in the number of lags does not provide any increase in R^2s. These results are not surprising given the fact that the models have a small number of independent parameters and exponential decay of the lag coefficients.

The results change in an environment with microstructure noise. Namely, they depend not only on microstructure noise variance, but also on its kurtosis. For $\kappa \leq 2$, the RV estimator (RV) performs better compared to RV_{AC_1} and the averaging over subsamples realized estimator $\overline{\mathrm{RV}}$ produces better results than RV_{TS} for all frequencies. Moreover, in absolute terms, when microstructure noise variance $\sigma^2 \geq 0.015$, all R^2s monotonically decrease with the increase in the subsampling frequency for a given set of model parameters.

The aforementioned finding can be explained by the fact that, for this range of microstructure noise variances, the optimal sampling in terms of R^2 becomes lower than five minutes. The main factor that impacts the performance of R^2 is the variance of the realized volatility estimators. The biases are captured by the constant terms of the regressions. Fig. 1 shows the optimal sampling frequencies that maximize R^2 of the estimators, conditional on specific values for noise variance for $\kappa = 3$. For large values of market microstructure noise variance increasing the number of lags is helpful. For some extreme cases the difference between the R^2 for a regression with 15 lags and a regression with 50 lags is 10%. The explanatory power of the group A estimators monotonically decreases for all values of microstructure noise variance considered.

For $\kappa > 2$ the situation changes. First, for higher frequencies, the RV estimator R^2_{pl} is smaller than $R^2_{AC_1}$. Second, R^2_{TS} becomes larger than R^2_{av} for some range of near-optimal frequencies around the maximum of R^2. This result supports the findings of Aït-Sahalia et al. (2005) who show that the two-scales estimator performs better with the optimal frequency compared to the "second best" averaging over subsamples estimator. However, for the non-optimal frequencies the averaging over subsamples estimator outperforms the two-scales one.

The eigenfunction approach also allows us to derive an approximately optimal frequency. However, analytical solutions can only be obtained for the simplest cases of the RV_{AC_1} and RV estimators. For the other two cases the solution can be found as a root of third and fourth power polynomials. For these two it is more

6 See, for example, Hansen and Lunde (2006) and Sinko (2007).

Table 1
Sample Theoretical R^2 Comparison of MIDAS approach for the M1–M3 models. Each entry in the table corresponds to the R^2 for the different models (A.24)–(A.26), different number of lags and the different return sampling frequencies. The regressions are run on a weekly (5 days) data sampling scheme. The names of the variables are consistent with the section describing realized volatility estimators. Every column in the panel corresponds to the theoretical explanatory power of the different left-hand side variables for the same right-hand side variable. The first panel contains theoretical results for the "noiseless" case, the second contains results for the case of *i.i.d.* noise with $\kappa = 3$ and $\sigma^2 = 0.001$, and the third contains the case of *i.i.d.* noise with $\kappa = 3$ and $\sigma^2 = 0.03$. The complete results are provided in Ghysels and Sinko (2009), Table B-1.

		1 lag					50 lags				
	R^2_{best}	RV_{IV}	RV	RV_{AC_1}	$\overline{RV}$	RV_{TS}	RV_{IV}	RV	RV_{AC_1}	$\overline{RV}$	RV_{TS}
LHS: IV, $\sigma^2 = 0.0000$											
$M1^{5\text{min}}$	**0.891**	**0.871**	0.799	0.686	0.822	0.817	**0.874**	0.822	0.776	0.834	0.830
$M1^{1\text{min}}$	**0.891**	**0.871**	0.856	0.827	0.861	0.832	**0.874**	0.856	0.836	0.861	0.839
$M1^{20\text{sec}}$	**0.891**	**0.871**	**0.866**	0.856	**0.867**	0.783	**0.874**	**0.866**	0.856	**0.868**	0.814
$M2^{5\text{min}}$	**0.586**	**0.445**	0.346	0.240	0.375	0.372	**0.460**	0.390	0.328	0.407	0.406
$M2^{1\text{min}}$	**0.586**	**0.445**	0.422	0.381	0.429	0.415	**0.460**	0.439	0.411	0.445	0.434
$M2^{20\text{sec}}$	**0.586**	**0.445**	**0.437**	0.422	**0.440**	0.396	**0.460**	**0.452**	0.439	**0.454**	0.421
$M3^{5\text{min}}$	**0.945**	**0.934**	0.875	0.775	0.894	0.888	**0.936**	0.900	0.869	0.908	0.905
$M3^{1\text{min}}$	**0.945**	**0.934**	0.922	0.898	**0.926**	0.895	**0.936**	0.923	0.910	**0.926**	0.908
$M3^{20\text{sec}}$	**0.945**	**0.934**	**0.930**	0.922	**0.932**	0.841	**0.936**	**0.930**	0.923	**0.932**	0.888
LHS: IV, $\sigma^2 = 0.0010$, $\kappa = 3.0$											
$M1^{5\text{min}}$	**0.891**	**0.871**	0.774	0.668	0.822	0.816	**0.874**	0.810	0.770	0.833	0.830
$M1^{1\text{min}}$	**0.891**	**0.871**	0.809	0.790	0.860	0.831	**0.874**	0.827	0.818	0.860	0.839
$M1^{20\text{sec}}$	**0.891**	**0.871**	0.778	0.789	**0.863**	0.779	**0.874**	0.812	0.817	0.863	0.813
$M2^{5\text{min}}$	**0.586**	**0.445**	0.301	0.218	0.375	0.372	**0.460**	0.364	0.314	0.407	0.406
$M2^{1\text{min}}$	**0.586**	**0.445**	0.321	0.309	0.427	0.413	**0.460**	0.375	0.368	0.443	0.432
$M2^{20\text{sec}}$	**0.586**	**0.445**	0.261	0.284	0.425	0.386	**0.460**	0.341	0.354	0.442	0.414
$M3^{5\text{min}}$	**0.945**	**0.934**	0.852	0.757	0.894	0.888	**0.936**	0.892	0.865	0.908	0.905
$M3^{1\text{min}}$	**0.945**	**0.934**	0.880	0.865	**0.925**	0.895	**0.936**	0.902	0.896	0.926	0.908
$M3^{20\text{sec}}$	**0.945**	**0.934**	0.848	0.860	**0.927**	0.837	**0.936**	0.891	0.895	**0.927**	0.887
LHS: IV, $\sigma^2 = 0.0300$, $\kappa = 3.0$											
$M1^{5\text{min}}$	**0.891**	**0.871**	0.123	0.153	0.807	0.806	**0.874**	0.471	0.508	0.826	0.825
$M1^{1\text{min}}$	**0.891**	**0.871**	0.032	0.046	0.603	0.640	**0.874**	0.253	0.309	0.747	0.760
$M1^{20\text{sec}}$	**0.891**	**0.871**	0.011	0.017	0.181	0.205	**0.874**	0.124	0.166	0.536	0.557
$M2^{5\text{min}}$	**0.586**	**0.445**	0.012	0.015	0.340	0.346	**0.460**	0.048	0.061	0.386	0.390
$M2^{1\text{min}}$	**0.586**	**0.445**	0.003	0.004	0.117	0.145	**0.460**	0.013	0.018	0.234	0.260
$M2^{20\text{sec}}$	**0.586**	**0.445**	0.001	0.001	0.018	0.021	**0.460**	0.004	0.007	0.068	0.078
$M3^{5\text{min}}$	**0.945**	**0.934**	0.148	0.185	0.879	0.877	**0.936**	0.624	0.657	0.902	0.901
$M3^{1\text{min}}$	**0.945**	**0.934**	0.039	0.056	0.670	0.706	**0.936**	0.401	0.465	0.842	0.852
$M3^{20\text{sec}}$	**0.945**	**0.934**	0.014	0.020	0.213	0.239	**0.936**	0.224	0.286	0.677	0.694

convenient to express the solution in terms of $\phi \equiv \overline{M}/M$, $\phi \in (0, 1)$ (Bandi and Russell, 2008a). As a result, $\phi \simeq 0$ corresponds to the case of the lowest possible frequency, $\phi \simeq 1$ corresponds to the case of the highest frequency. In our analysis only $\overline{M}$ changes. There are two major differences between the optimal sampling we derive and the optimal sampling derived in Bandi and Russell (2008b). First, for prediction purposes, the bias of the estimator does not matter as long as it is constant over time. As a result, using Theorem 2.2, the optimal sampling frequency is the one that minimizes the variance of the estimators. Second, we are using a stochastic volatility model while Bandi and Russell (2008b) assume that the variance is deterministic function of time. The following proposition summarizes our results:

Proposition 2.1. *Let Assumption 2.1 and the conditions of Theorem 2.2 hold, $\overline{M}h \simeq 1$, $m \simeq M/\overline{M}$. Also, lets define $Q = \sum_{i=0}^{p} a_i^2$. Then the approximate optimal sampling frequencies for the* RV, RV_{AC1}, RV_{TS} *and* $\overline{RV}$ *estimators are:*

$$\overline{M}_{\text{RV}} \simeq \sqrt{\frac{Q}{2\kappa\sigma_\eta^4}}, \qquad \overline{M}_{\text{RV}_{AC_1}} \simeq \sqrt{\frac{3Q}{4\sigma_\eta^4}}$$

$$\overline{M}_{\overline{\text{RV}}} \simeq \arg\min\left\{8a_0\phi\sigma_\eta^2 + 4\phi^2 M\kappa\sigma_\eta^4 - 2\phi\sigma_\eta^4(\kappa-1) + \frac{Q(1-\phi)\{2M(2-\phi)-(1/\phi+1)\}}{3M^2\phi}\right\}$$

$$\overline{M}_{\text{RV}_{TS}} \simeq \arg\min\left\{\frac{Q\{2M(2-\phi)-(1/\phi+1)\}}{3M^2\phi(1-\phi)} - \frac{2\phi\sigma_\eta^4(\kappa-1)}{1-\phi} + \frac{8\phi a_0\sigma_\eta^2 + 8\phi^2 M\sigma_\eta^4}{(1-\phi)^2}\right\}. \qquad (2.10)$$

Proof. See Appendix A.3. □

We note from the above proposition that $\overline{M}_{\text{RV}}/\overline{M}_{\text{RV}_{AC_1}} \simeq \sqrt{2/(3\kappa)}$. Although the result is based on approximations, it matches with exact computations. In particular, consider the findings reported in Fig. 1 where we computed the R^2 as a function of sampling frequency, second and fourth moments of market microstructure noise and mean and variance of realized volatility directly for three models M1–M3 ((A.24)–(A.26)) using an exact formula for the variance. The optimal sampling frequency for RV_{AC1} estimator is always greater than the optimal sampling frequency for RV estimator, and the difference increases as kurtosis κ of market microstructure noise increases.

We showed in Theorem 2.2 that the optimum in terms of MSE of prediction is equivalent to the optimum in terms of minimum of estimator's variance. Comparing optimal frequencies in terms of *MSE of prediction* and optimal frequency in terms of *MSE of estimator* we can conclude that they will coincide whenever the estimators are unbiased. For cases when the bias is non-zero and is linearly increasing with the sampling frequency, the optimal sampling frequency in terms of MSE of prediction will be higher

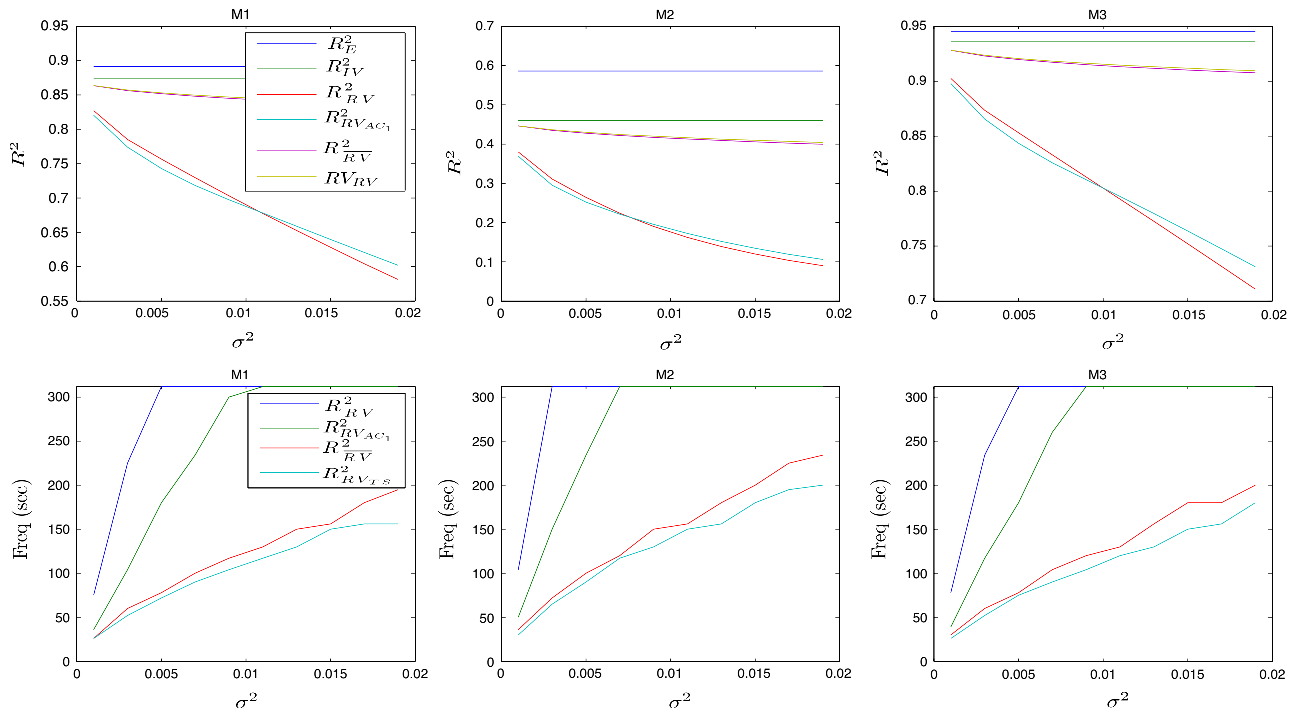

Fig. 1. Optimal sampling, $\kappa = 3$. Optimal sampling frequency and maximal R^2 of the different models and different realized volatility estimators. R^2_E corresponds to the prediction power of the conditional expectation, R^2_{IV} corresponds to the prediction power of the integrated volatility.

than the optimal sampling frequency in terms of MSE of estimator, as the increase in the bias introduces an additional penalty as the sampling frequency increases. Also, note that Bias(RV) $=$ Bias($\overline{\text{RV}}$) $= 2\overline{M}\sigma^2_\eta$.

3. Practical implementation issues

In this section we discuss various practical implementation issues, ranging from the choice of regression models, the data and optimal sampling schemes.

We compare in this subsection realized volatility estimators using the forecast performance of MIDAS regressions proposed, in the context of volatility prediction, by Ghysels et al. (2006). We will consider the beta polynomial specification for the parameterization.

3.1. Data

We use the exact same prefiltered price data as in Hansen and Lunde (2006). The data is provided by the authors. The data consists of the thirty equities of the Dow Jones Industrial Average (DJIA). The sample period spans five years, from January 3, 2000 to December 31, 2004. All the data are extracted from the Trade and Quote (TAQ) database. In particular, we use the trade prices for our analysis. The raw data were filtered for outliers and transactions outside the period from 9:30 am to 4:00 pm were discarded. The filtering procedure removed obvious data errors such as zero prices.

For most of our estimators we use calendar-time sampling. It requires the construction of artificial prices from the original tick-by-tick irregularly-spaced price data. For interpolation we use the previous-tick method, introduced by Wasserfallen and Zimmermann (1985).

3.2. Unconditional and conditional optimal sampling

Optimal sampling frequency issues were first considered, for the homoscedastic case, by Zhou (1996). The idea was further developed by Oomen (2005), Bandi and Russell (2008a) and Hansen and Lunde (2006) among others. Following Bandi and Russell (2006) we use term "conditional" to reflect the fact that the optimal sampling frequency for realized volatility estimators is computed on a daily basis using Proposition 2.1 formulas with daily estimates of second and fourth moments of market microstructure noise and quarticity. In contrast, unconditional optimal sampling fixes the sampling frequency over the whole period. In the remainder of this subsection we will discuss the microstructure noise moment estimators used to compute the conditional optimal sampling frequencies.

We are partially adopting the Bandi and Russell approach, though our focus is to find optimal sampling in terms of *MSE in prediction*, not *MSE in estimator*. To apply the results of Proposition 2.1, we need to estimate the second and fourth moments of market microstructure noise as well as a measure of daily quarticity. The quarticity and the fourth moment of market microstructure noise can be estimated from sum of returns in fourth power, R^m_j, using different sampling frequencies (note that the number of observations on a grid M^m_j is $\overline{M}+1$, $h \simeq 1/\overline{M}$):

$$\mathrm{E}\left(R^m_j\right) \equiv \mathrm{E}\left(\sum_{t_{k,-1}\in\mathcal{M}^m_j} r^4_{t_k,h}\right) \underset{h\ll 1}{\simeq} 3hQ + 12a_0\sigma^2_\eta + \frac{2}{h}\left\{\mathrm{E}\left(\eta^4\right) + 3\mathrm{E}\left(\eta^2\right)^2\right\}. \quad (3.1)$$

In the no-noise (relatively low-frequency) environment R^m_j is used to compute Barndorff-Nielsen and Shephard (2002) quarticity estimator $\hat{Q}^m_j$, which can be further improved using subsampling averaging. In the noisy environment (the highest-frequency case) it is used to estimate $\widehat{\mathrm{E}\left(\eta^4\right)}$.

$$\hat{Q} = \frac{1}{m}\sum_j^m \hat{Q}^m_j = \frac{1}{m}\sum_j^m \frac{\overline{M}}{3}R^m_j; \qquad \widehat{\mathrm{E}\left(\eta^4\right)} \simeq \frac{1}{2M}R^1_1 - 3\hat{\sigma}^4_\eta. \quad (3.2)$$

To measure realized variance we use Eq. (A.5) at low frequency. To measure second moment of microstructure noise we use estimator proposed by Bandi and Russell (2006):

$$\hat{\sigma}_\eta^2 \simeq \frac{1}{2M} \sum_{t_{k,-1} \in \mathcal{M}} (r_{t_k})^2 \tag{3.3}$$

where $\mathcal{M}$ is the finest possible grid (every second), $M = 23\,400$ is the number of elements in the finest possible grid minus one, $\mathcal{M}_j^m$ is fifteen minutes grid, m is a number of grids.

4. Empirical results

We are armed with theoretical predictions about how various volatility measures should behave as far as prediction is concerned. How does this all play out in real data? In this section we describe both the in-sample and out-of-sample empirical forecasting performance for the different estimators. Due to space limitations we only present the details of the optimal sampling results. Before we do we briefly describe the one-minute and five-minute frequency results that can be found in a technical appendix (Ghysels and Sinko, 2009). We find that, within the MIDAS framework, the five-minutes frequency can be considered as a low-noise frequency environment where all of our theoretical predictions hold. In contrast, for the one-minute frequency it seems that we do not find many coherent empirical results, nor results that square with the theory. In general, the out-of-sample results support our in-sample prediction findings. Averaging over subsamples estimators based on five-minutes returns outperform the rest. RV_{TS} and $\overline{\mathrm{RV}}$ estimators perform the same. Comparing the RV estimator performance with that of two scales and averaging over subsamples reveals that the RV estimator produces worse results. Finally, one-minute out-of-sample and in-sample predictions share similar patterns. There is no uniformly better estimator in terms of predicting power for this sampling frequency. In the remainder of this section, we focus on optimal sampling in terms of MSE of prediction.

There are at least two reasons why we want to examine empirically optimal sampling issues in the context of prediction of volatility in the presence of microstructure noise. First, we noted so far that the five-minute empirical results aligned with the theoretical predictions regarding realized volatility measures, yet for the one-minute frequency there is no volatility measure that uniformly outperforms the others. Second, in Section 2.3 we established a number of theoretical results that we now verify empirically. Notably, we showed in Section 2.3 that the optimal sampling in terms of MSE of prediction should be higher than those in terms of MSE of the estimator.

This section answers the following questions: (i) What is the unconditional optimal frequency for individual stocks? (ii) What is the conditional optimal frequency implied by estimated daily characteristics of market microstructure noise and realized volatility? (iii) What is the gain in terms of R^2 for the second method compared to the first? To do this, we need to estimate daily variance and quarticity of efficient log-price process as well as the second and fourth moment of market microstructure noise.

When considering optimal sampling we need to broaden the empirical specification used for the regressions, as the sampling frequency of LHS and RHS variables will differ. In particular, our analysis makes use of the following regression framework

$$\mathrm{RV}_y^m(t+H,t) = \mu_H^{m'} + \phi_H^{m'} \sum_{k=0}^{k^{\max}} b_H^{m'} \times (k,\theta) \mathrm{RV}_{y'}^{m'}(t-k,t-k-1) + \epsilon_{Ht}^{m'}, \tag{4.1}$$

where for the left hand side (1) $\mathrm{RV}_y^m = \mathrm{RV}_{TS}$ is computed using five-minute, one-minute or two-second sampling frequencies m and (2) H is a five-day period. For the right hand side we use (1) a set of variables $\mathrm{RV}_{y'} = \{\mathrm{RV}, \mathrm{RV}_{TSd}, \mathrm{RV}_{AC_1}, \mathrm{RV}_{NW}, \overline{\mathrm{RV}}, \mathrm{RV}_{TS}\}$ constructed using sampling frequencies m' from 2 s to 10 min and (2) the conditional optimal sampling frequency determined empirical using the daily estimators from Section 3.2.

To keep the empirical analysis concise we study two stocks: IBM and AA as representatives of a liquid and a relatively illiquid stock. The results for the unconditional and conditional sampling frequencies are presented in Figs. 2 and 3. Each stock has six subplots corresponding to the different $\mathrm{RV}_{y'}^{m'}$. For the construction of each plot we used the entire 2000–2004 sample. Our findings do not change if use only subsamples 2000–2002 or 2003–2004. The R^2's obtained with the conditional optimal sampling, which vary every day, are plotted using colors associated with the sampling frequency of the LHS variables. We use red for the 2 s frequency, green for the 1 min and blue for 5 min. The vertical lines correspond to *ex ante* unconditional sampling frequencies based on the unconditional measures of the noise and signal moments and Proposition 2.1.

Before discussion the empirical findings, let us first recall what the theory should tell us about the patterns of predictive power as we change the sampling frequency in the measurement of volatility. As mentioned in Section 2.3, we divide the estimators into two groups. Group $A = \{\mathrm{RV}, \mathrm{RV}_{AC_1}\}$ consists of estimators that have increasing variances as a function of the sampling frequency, whereas Group $B = \{\overline{\mathrm{RV}}, \mathrm{RV}_{TS}\}$ consists of the estimators with variances that decrease as the sampling frequency increases. Hence, one might expect from theory that Group B estimators should outperform Group A estimators for the relatively high frequencies. However, this statement is based on asymptotic arguments, which may be or may not be a good description of what we see in empirical applications. Indeed, the decreases of the variance in the Group B cases, is based on asymptotic arguments that may not directly apply to our empirical analysis since the finest possible grid M is constant, and the asymptotics assume that as the subsampling frequency increases, the highest frequency associated with M increases, too. Therefore, due to the fact that lower frequencies group B estimators have lower discretization error, the explanatory power of group B estimators for lower frequencies should be higher compared to the explanatory power of group A estimators. By the same reason, the optimal unconditional frequency for these estimators should be higher compared to the RV estimator.

Along the same lines, what do we expect from theory for the Group $A = \{\mathrm{RV}, \mathrm{RV}_{AC_1}\}$ estimators? Let's start with the RV estimator. As the sampling frequency increases, the discretization noise of the estimator decreases. At the same time, the impact of the market microstructure noise becomes more significant. Thus the R^2 as a function of the sampling frequency m (or the number of log-price observations per day $\overline{M}$) can be either increasing (the sampling frequency is too low to achieve the optimum), decreasing (the sampling frequency is too high to achieve the optimum), or hump-shaped (there is an optimum within the interval considered). The same pattern should hold for the Zhou estimator RV_{AC_1}. Moreover, as sampling frequency increases, the probability of two consecutive non-zero returns decreases. Combining with the previous-tick method we use, this leads to the convergence of the RV_{AC1} estimator for the ultra-high frequencies to the RV estimator. In addition, the maximum of the "hump" for RV_{AC1} should be achieved at a higher frequency than for the "plain vanilla" estimator (see discussion after Proposition 2.1).

The behavior of group B estimators depends not only on $\overline{M}$ but also on the finest possible grid over the day, i.e. M. It follows

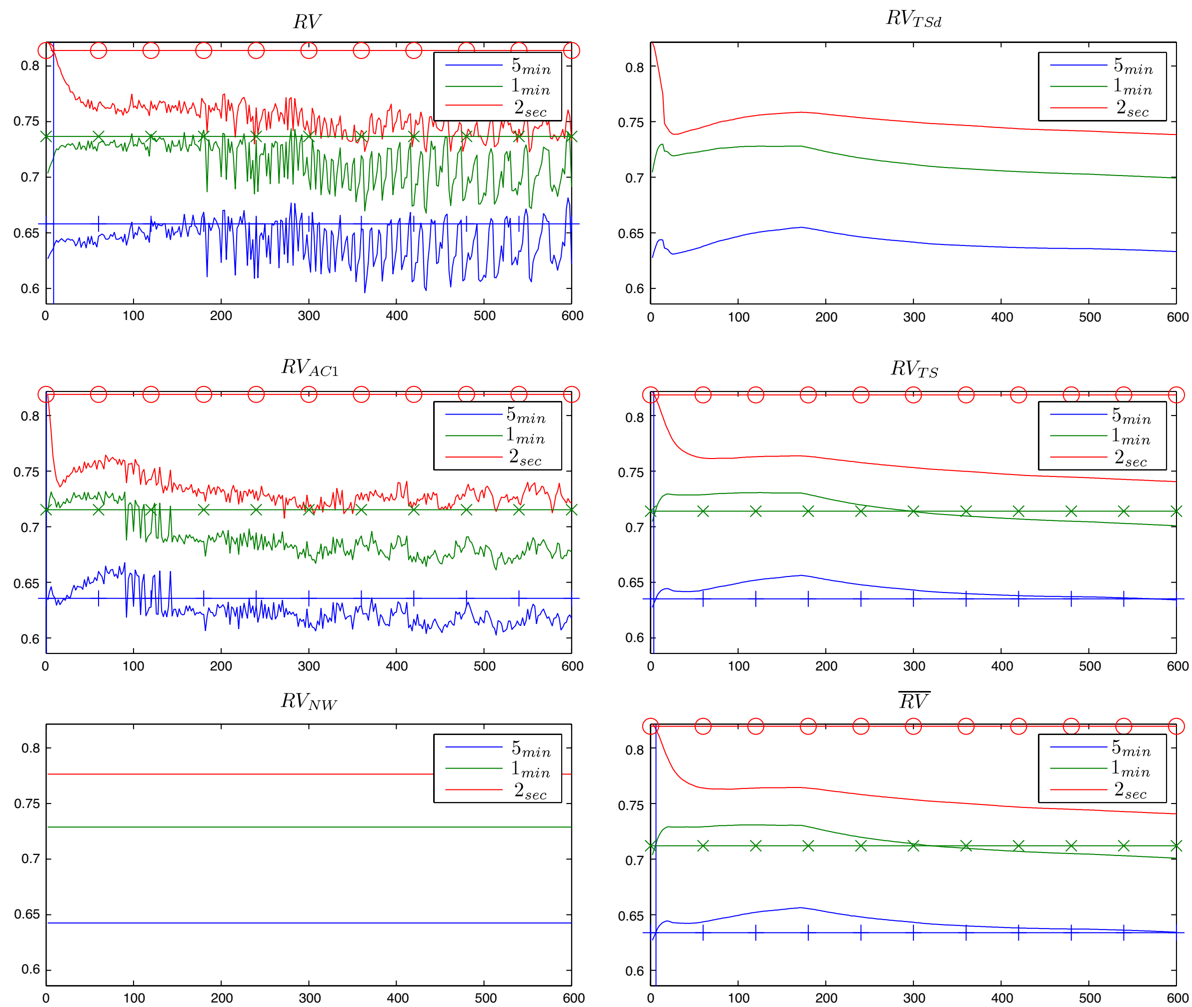

Fig. 2. R^2 as a function of frequency. IBM stock. Full sample. The figure shows dependence of regression R^2 as a function of sampling frequency. LHS variable is $\overline{RV}$, constructed using 5 min, 1 min and 2 s sampling frequencies. R^2 is computed for the following RHS: RV, RV_{AC1}, $\overline{RV}$, RV_{NW}, RV_{TS}, RV_{TSd} and from 2 to 600 s sampling frequencies. The explanatory power of the conditional optimal frequency for a given estimator is plotted using the same color as the unconditional one. The results are given for five-year sample (Jan. 2000–Dec. 2004).

directly from the proof of Proposition 2.1 that when the ratio $\phi = \overline{M}/M$ is large enough, the $\overline{\text{RV}}$ estimator behaves like the "plain vanilla" realized volatility estimator. The behavior of the variance of RV_{TS} resembles that of $\overline{\text{RV}}$.

Further, we would expect that the change in the LHS sampling frequency will increase the explanatory power of the regression if the resulting error (sum of the discretization and the market microstructure noise) decreases. That is, as long as the decrease in the discretization noise does not compensate for the increase in the market microstructure noise, the explanatory power of the regression increases, and vice versa. In addition, the unconditional sampling frequency optimum (the maximum of the hump) should stay constant.

How do these plots match up with the theoretical predictions? As we mentioned in Section 2.3, the *i.i.d.* market microstructure noise assumption is most questionable. As a result, empirical violation of the theory is most likely due to the time-varying market microstructure noise. There are some patterns that contradict the theory. namely: (a) the patterns of the plots should look like parallel shifts across the three colors, and this not the case for AA and it also not the case for the high frequency patterns of the red lines for IBM, (b) conditional optimal sampling should be at least as good as unconditional optimal sampling. Conditional optimal sampling yields predictions that are reasonably close to the maximal R^2 of the regressions for each fixed frequency. Yet, in general they are slightly worse than the unconditional ones, and (c) the red lines are upward trending for all estimators, towards the high frequencies, and this seems to imply that the variance of microstructure noise is predictable.

Nevertheless, some findings square with the theory. All these findings are related to the lower-frequencies and stocks with higher liquidity, namely: (a) Group B estimators outperform Group A estimators for the relatively low frequencies. This confirms the simulation results reported in Aït-Sahalia and Mancini (2008), (b) comparing the Zhou estimator with all other estimators reveals that it performs poorly for low frequencies, confirming earlier results reported in Ghysels and Sinko (2006) and Ghysels et al. (2007), (c) the unconditional optimal sampling frequency of the Zhou estimator is higher than the one for the RV estimator, (d) the patterns of the plots looks like parallel shifts across the blue and green colors in case for IBM stock and for the lower frequency segment of AA stock. (e) the pattern of of the $\overline{\text{RV}}$ estimator R^2 is roughly identical to the prediction power of the RV_{TS} estimator.

In the remainder of the section we take a closer look at respectively the contradictions and the support for the theoretical findings. It appears that the contradictions occur mostly in the cases where the LHS and RHS variables contain a significant amount of market microstructure noise. For example, for IBM stock

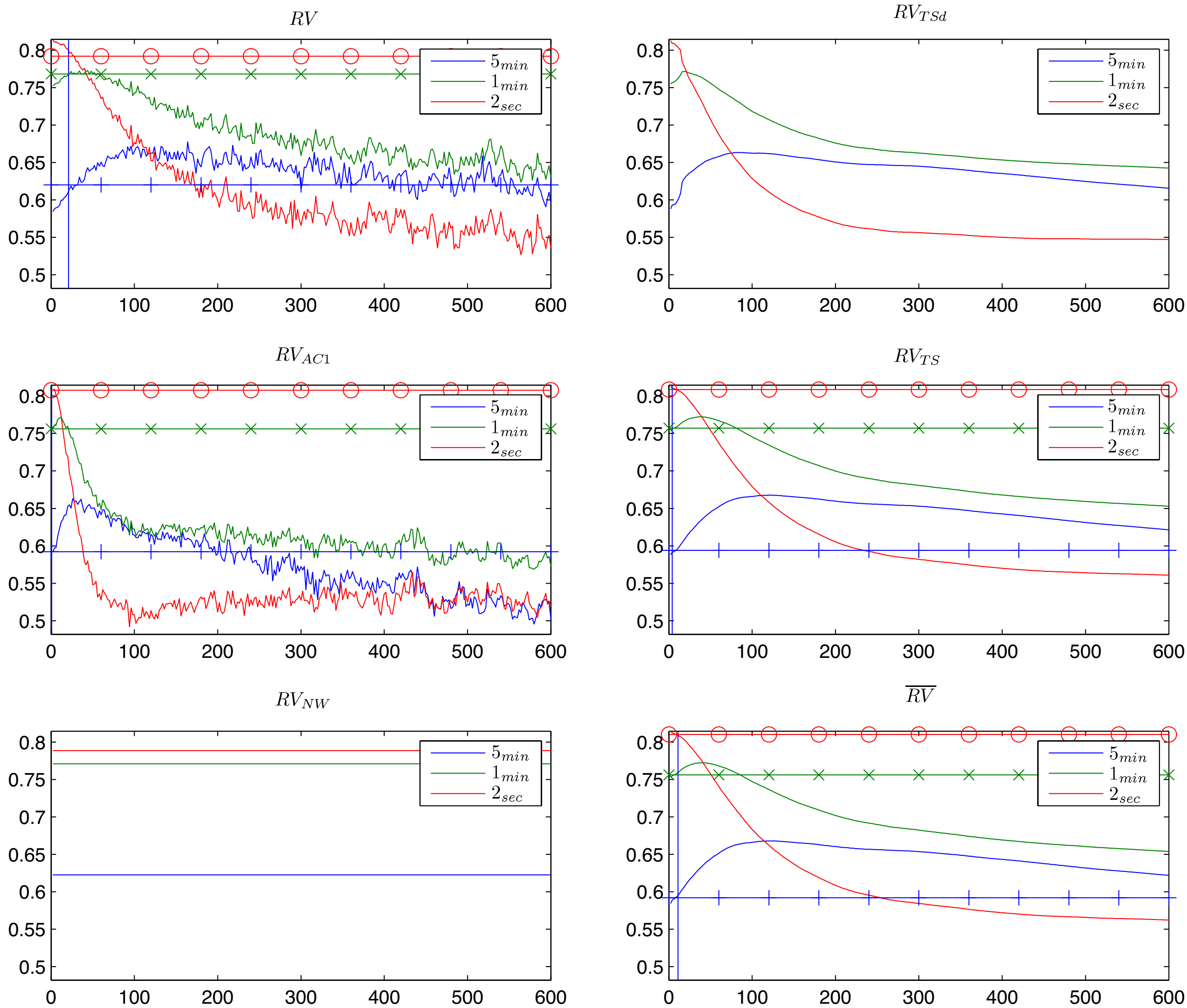

Fig. 3. R^2 as a function of frequency. AA stock. Full sample. The figure shows dependence of regression R^2 as a function of sampling frequency. LHS variable is $\overline{RV}$, constructed using 5 min, 1 min and 2 s sampling frequencies. R^2 is computed for the following RHS: RV, RV_{AC1}, $\overline{RV}$, RV_{NW}, RV_{TS}, RV_{TSd} and from 2 to 600 s sampling frequencies. The explanatory power of the conditional optimal frequency for a given estimator is plotted using the same color as the unconditional one. The results are given for five-year sample (Jan. 2000–Dec. 2004).

the theory holds for the five-minute and one-minute LHS sampling frequency and for the entire range of the RHS frequencies. The largest number of discrepancies from the theory can be observed for the two-second LHS and the RHS, which is frequent enough to contain significant amount of the market microstructure noise. This seems to suggest that, particularly for an illiquid stock like AA, the variance of microstructure noise is predictable, a topic of further research covered in Sinko (2007).

We noted that, in support of the theory, group B estimators have the same predictive power patterns. These two estimators have asymptotically zero variance, and the two-scales correction only removes the bias from the $\overline{RV}$ estimator. However, if the bias is constant over time, it does not affect the predictive power of the regression as it only changes the intercept. Therefore, the performance of these estimators should be approximately the same, and this is exactly what we observe in the data. Note however that, as a caution, in real-data applications with possibly time-varying variance of market microstructure noise, the RV_{TS} should be preferred.

Finally, we would like to discuss the conditional optimal sampling results. They are reasonably close to the maximal R^2 of the regressions for each frequency, even though they are in general worse than the unconditional ones. To estimate the optimal frequencies, we use the approximation derived in Proposition 2.1 using parameter estimates from Section 3.2.

The histograms of kurtoses and conditional optimal sampling frequencies implied by the theoretical model for the AA and IBM stocks appear in Fig. 4. The major differences in the conditional optimal sampling frequencies across the two stocks can be captured by the difference of the kurtoses between them. The kurtosis histogram of the AA stock is much wider compared to the same histogram of the IBM stock. This is the main factor that widens the conditional sampling frequency histograms. For the optimal frequency of the Zhou estimator, which does not depend on the market microstructure kurtosis, the conditional optimal sampling frequency is usually higher than the highest sampling frequency considered (2 s). The histograms of this estimator looks the same for the two stocks considered. The model implies that the optimal sampling frequency for the RV_{TS} estimator is higher than the optimal sampling frequency of the $\overline{RV}$ estimators, even though it does not change the explanatory power of the regression in a significant way. The RV_{AC1} optimal sampling frequency is much higher than the optimal frequencies of the other estimators. As a result, for the lower frequencies

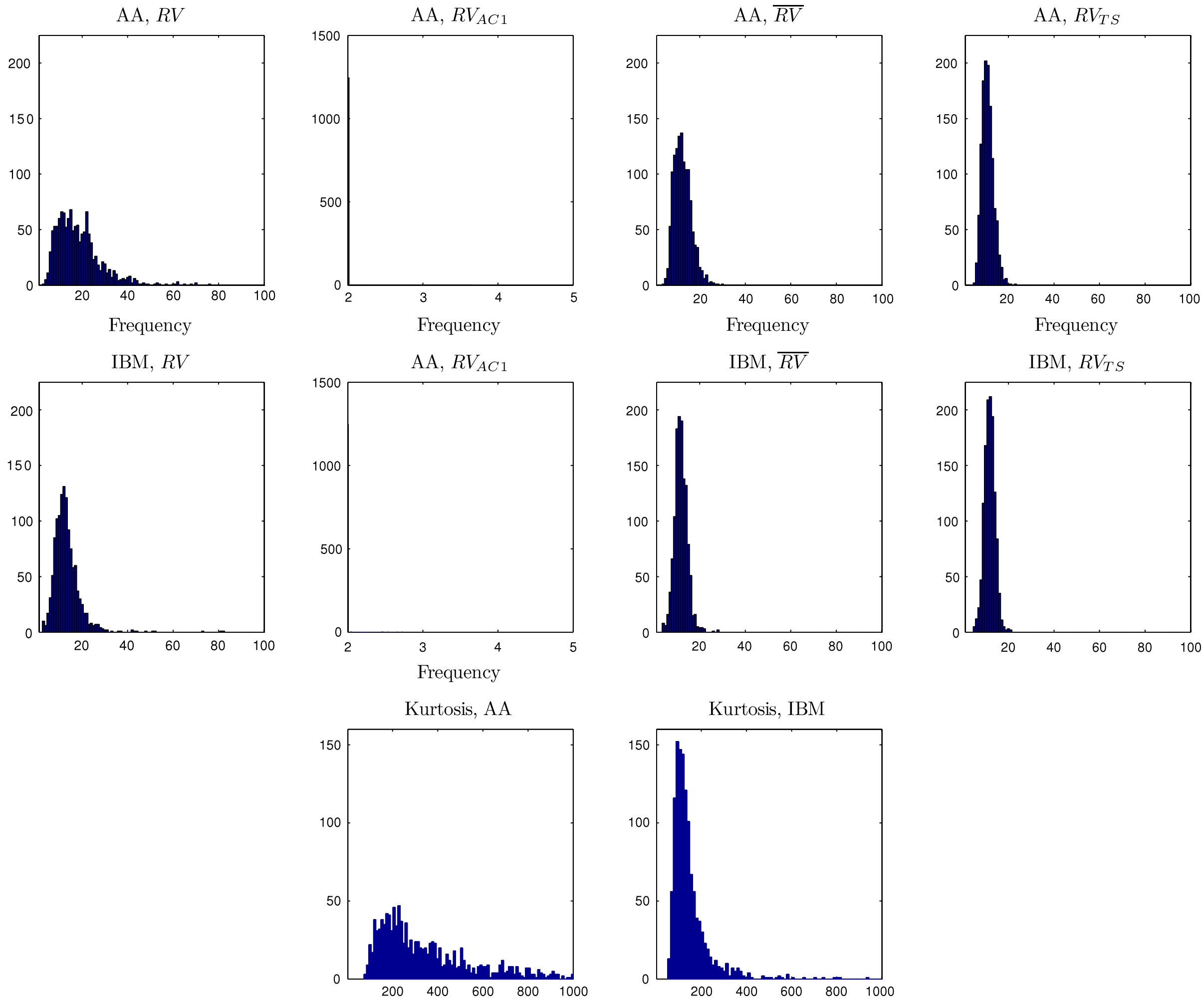

Fig. 4. Estimated kurtosis and conditional sampling frequencies for AA and IBM stocks. The figure shows estimated kurtosis and approximate conditional optimal sampling frequencies constructed on a daily basis using Proposition 2.1 and Section 3.2. The results are provided for IBM and AA stocks and five-year sample (Jan. 2000–Dec. 2004).

(5 min, 1 min) it is usually outperformed by the others. The RV estimator has a higher optimal sampling frequency compared to the RV_{TS} and $\overline{RV}$ estimators. As mentioned before, this is a result from the fact that in this subsection the finest possible grid (1 s frequency) stays constant as the subsampling frequency increases.

5. Conclusions

We studied the forecasting of future volatility using past volatility measures unadjusted and adjusted for microstructure noise. We examined the population properties of a regression prediction problem involving measures volatility that is contaminated by noise. The general regression framework allowed us to compare the population performance of various estimators and also study of optimal sampling issues in the context of volatility prediction with microstructure noise. We also conducted an extensive empirical study of forecasting with microstructure noise using the 30 Dow Jones stocks. Our empirical results suggest that for this data, within the class of quadratic variation measures, the subsampling and averaging approach (see Zhang et al. (2005)) constitutes the class of estimators that performs best in a prediction context. Overall our empirical findings with five minute sampling schemes square with our theory developed in the paper and confirm earlier findings reported Aït-Sahalia and Mancini (2008), Ghysels and Sinko (2006) and Ghysels et al. (2007). We also examined the empirics of optimal conditional and unconditional sampling. The optimal sampling exercise compares the explanatory power patterns implied by the theory with the ones estimated from the data. This comparison demonstrates that the theory provides a reasonable explanation for many features of the empirical data for a liquid stock like IBM. For an illiquid stock, like AA, the findings do not square as much with the theory. We conjecture that what is missing, is a model that can capture the more complex time-dependent characteristics of market microstructure noise. This is further explored in Sinko (2007).

Appendix. Technical details

A.1. Volatility estimators

Define two time grids. $\mathcal{M} = \{t_0, \ldots, t_M\}$ corresponds to the largest possible number of equally-spaced observations per day measured in seconds. $\mathcal{T} = \{\tau_0, \ldots, \tau_T\}$ corresponds to the actual-time records of the transaction tick-by-tick price data. Any equidistant subsample grid can be represented as $\mathcal{M}_j^m = \{t_j, t_{j+m}, t_{j+2m}, \ldots\}$ with $j = 0, \ldots, m-1$, $\bigcup_{j=0,\ldots,m-1} \mathcal{M}_j^m = \mathcal{M}$, and $\bigcap_{\forall i \neq j} \mathcal{M}_j^m \mathcal{M}_i^m = \emptyset$. Log-returns on the grid $\mathcal{M}^m$ for two consecutive times $t_{k-1}, t_k \in \mathcal{M}^m$ and Eq. (2.1) is defined as

$$r_{t_k,m} = p_{t_k} - p_{t_{k-1}} = p^*_{t_k} - p^*_{t_{k-1}} + \eta_t - \eta_{t_{k-1}} = r^*_{t_k,m} + e_{t_k,m}. \quad \text{(A.1)}$$

Under Assumption 2.1 $\mathrm{Var}(e_t^m) = 2\sigma_\eta^2$. The first estimator we consider is the daily RV^m estimator:

$$\begin{aligned}\mathrm{RV}^m &= \sum_{t_k,t_{k-1}\in\mathcal{M}^m}(r_{t_k,m})^2\\ &= \sum_{t_{k,-1}\in\mathcal{M}^m}(r^*_{t_k,m})^2 + 2\sum_{t_{k,-1}\in\mathcal{M}^m} e_{t_k,m}r^*_{t_k,m} + \sum_{t_{k,-1}\in\mathcal{M}^m} e^2_{t_k,m}.\end{aligned} \tag{A.2}$$

To simplify notations, we use $t_{k,-1} \in \mathcal{M}^m$ instead of $t_k, t_{k-1} \in \mathcal{M}^m$. Without market microstructure noise ($e = 0$) the previous equation defines a consistent estimator of integrated variance with respect to the number of observations $\overline{M}$ in the subsample (sampling frequency). However, it is biased and inconsistent in the presence of microstructure noise.

The second estimator we consider is studied by Zhou (1996) and Hansen and Lunde (2006). We adopt the notation of the authors, and call it the $\mathrm{RV}^m_{AC_1}$ estimator:

$$\mathrm{RV}^m_{AC_1} = \tilde{\gamma}_0^m + 2\tilde{\gamma}_1^m, \qquad \tilde{\gamma}_j^m = \frac{\overline{M}}{\overline{M}-j}\sum_{t_{k,-1}\in\mathcal{M}^m} r^m_{t_k} r^m_{t_{k+j}}. \tag{A.3}$$

Under *i.i.d.* assumption of the microstructure noise, $E(r^m_{t_{k+1}} r^m_{t_k}) = -\sigma^2_\eta$ and $E(r^2_{t_{im}}) = \mathrm{Var}(r^*_{t_{im}}) + 2\sigma^2_\eta$, and thus it is unbiased. The third estimator is consistent. It is proposed by Aït-Sahalia et al. (2005)

$$\begin{aligned}\mathrm{RV}^m_{TS} &= \frac{1}{m}\sum_{j=0}^{m-1}\sum_{t_{k,-1}\in\mathcal{M}^m_j}(r^m_{t_k})^2 - \frac{\overline{M}}{M}\sum_{t_{k,-1}\in\mathcal{M}}(p_{t_{k+1}} - p_{t_k})^2\\ &= \frac{1}{m}\sum_{j=0}^{m-1}\mathrm{RV}^{m,j} - \frac{\overline{M}}{M}\mathrm{RV}^1,\end{aligned} \tag{A.4}$$

where $\mathrm{RV}^{m,j}$ is realized volatility associated with subgrid $\mathcal{M}^m_j$, number of observations per day starting from the jth observation is $\overline{M}$, and RV^1 is the realized volatility computed using all equally-spaced data available.

$$\overline{\mathrm{RV}}^m = \frac{1}{m}\sum_{j=0}^{m-1}\sum_{t_{k,-1}\in\mathcal{M}^m_j}(r^m_{t_k})^2. \tag{A.5}$$

The last two estimators we consider capture the fact that, in reality, microstructure noise can be serially correlated. The first estimator is a modification of the estimator defined at (A.4) and introduced in Aït-Sahalia et al. (2005). Instead of zero correlation of the noise component, they use a much weaker restriction:

$$\mathrm{Corr}(e_{\tau_i}, e_{\tau_{i+k}}) \le \rho^k, \quad \text{for some } |\rho| < 1, \tag{A.6}$$

where $\tau_1 < \cdots < \tau_T$ are the sequence of transaction tick-by-tick times that belong to $\mathcal{T}$. Under these conditions, the RV_{TSd} estimator is unbiased. Define the minimum step corresponding to "near uncorrelated frequency" as m' and the associated sample size of the subsample as $\overline{M}'$. The modified estimator RV_{TSd} is

$$\mathrm{RV}_{TSd} = \frac{1}{m}\sum_{j=0}^{m-1}\mathrm{RV}^{m,j} - \frac{\overline{M}}{\bar{M}'}\sum_{j=0}^{m'-1}\mathrm{RV}^{m',j}. \tag{A.7}$$

This equation is the analog of Eq. (A.4). The only difference is the second term that captures the fact of non-zero autocorrelation of the noise.

The last estimator is based on the tick-time grid $\mathcal{T}$ instead of the calendar-time grid $\mathcal{M}$. This estimator is proposed by Hansen and Lunde (2006). For T transactions occurring during the day at times τ_i and window w, based on the data sample, they conclude that it is enough to have about 15 lags for the noise to be "approximately uncorrelated". We name the last estimator $\mathrm{RV}^{1\ \mathrm{tick}}_{ACNW_w}$ and define it as

$$\mathrm{RV}^{1\ \mathrm{tick}}_{ACNW_w} = \sum_{j=-w}^{w}\frac{w-j}{w}\tilde{\gamma}_j^{1\ \mathrm{tick}} \tag{A.8}$$

where $\tilde{\gamma}_j^{1\ \mathrm{tick}} = \tilde{\gamma}_{|j|}^{1\ \mathrm{tick}} = T/(T-|j|)\sum_{i=1+|j|}^{T}(p_{t+\tau_i} - p_{t+\tau_{i-1}})(p_{t+\tau_{i-|j|}} - p_{t+\tau_{i-|j|-1}})$. Although this estimator is inconsistent, it provides material for comparison between tick-time and calendar time estimators for prediction purposes.

A.2. Proof of Theorems 2.1 and 2.2

Using properties of f_t defined in Assumption 2.1, properties for integrated variance $IV_{t,h}$ and square returns $(r^*_{t,h})^2$ over period h under no leverage assumption are: $IV_{t,h} = \int_{t-h}^{h}\sigma_t^2 dt$; $\sum_{i=1}^{N} IV_{t+ih,h} = \int_t^{t+Nh}\sigma_t^2 dt$, $\mathrm{E}(IV_{t+h,h}) = \int_t^{t+h} a_0 dt = a_0 h$; $\mathrm{Var}(IV_{t+h,h}) = 2\sum_{i=1}^{p}\frac{a_i^2}{\lambda_i^2}(\mathrm{e}^{-\lambda_i h} - 1 + \lambda_i h)$; $\mathrm{Cov}(IV_{t+h,h}, IV_{t-s,m}) = \sum_{i=1}^{p}\frac{a_i^2}{\lambda_i^2}\mathrm{e}^{-\lambda_i s}(1-\mathrm{e}^{-\lambda_i h})(1-\mathrm{e}^{-\lambda_i m})$, $(r^*_{t,h})^2 = \int_{t-h}^{t}\sigma_t^2 dt + 2\int_{t-h}^{t}\int_{t-h}^{u} dp_s dp_u = IV_{t,h} + Z_{t,h}$; properties of discretization error $Z_{t,h}$ under no leverage assumption, $\forall s, \delta \ge 0$:

$$\begin{aligned}&E(Z_{t,h}) = 0, \qquad \mathrm{Cov}(IV_{t,h}, Z_{t,h}) = 0,\\ &\mathrm{Var}(Z_{t,h}) = 4\left(\frac{a_0^2h^2}{2} + \sum_{i=1}^{p}\frac{a_i^2}{\lambda_i^2}(\mathrm{e}^{-\lambda_i h} - 1 + \lambda_i h)\right);\\ &E(Z_{t,h}Z_{t+\delta+s,s}) = 0.\end{aligned} \tag{A.9}$$

Lemma A.1. $\forall l > 0, \delta > 0, l > \delta$: $\mathrm{Cov}(Z_{t,s}, Z_{t+\delta,l}) = \mathrm{Var}(Z_{t,(l-\delta)\wedge s})$, $\delta \ge l$: $\mathrm{Cov}(Z_{t,s}, Z_{t+\delta,l}) = 0$.

Proof. For $l < s$:

$$\begin{aligned}Z_{t,s} &= 2\int_{t-s}^{t}\int_{t-s}^{u}\sigma_\tau dW_\tau \sigma_u dW_u\\ &\equiv 2\int_{t-s}^{t}\int_{t-s}^{u} dp_\tau dp_u\\ &= 2\int_{t-s}^{t}\left[\int_{t-s}^{u\wedge(t-l)} dp_\tau + \int_{t-l}^{u\vee(t-l)} dp_\tau\right]dp_u\\ &= 2\int_{t-s}^{t-l}\int_{t-s}^{u} dp_\tau dp_u\\ &\quad + 2\int_{t-l}^{t}\int_{t-s}^{t-l} dp_\tau dp_u + 2\int_{t-l}^{t}\int_{t-l}^{u} dp_\tau dp_u\\ &= Z_{t-l,s-l} + Z_{t,l} + 2r^*{}_{t-l,s-l}r^*{}_{t,l}\end{aligned} \tag{A.10}$$

with $\mathrm{Cov}(Z_{t-l,s-l}, 2r^*{}_{t-l,s-l}r^*{}_{t,l}) = \mathrm{Cov}(Z_{t,l}, 2r^*{}_{t-l,s-l}r^*{}_{t,l}) = 0$. For $\delta \ge l$,

$$\begin{aligned}\mathrm{Cov}(Z_{t,s}, Z_{t+\delta,l}) &= 4E\left(\int_{t-s}^{t}\int_{t-s}^{u} dp_\tau dp_u \int_{t+\delta-l}^{t+\delta}\int_{t+\delta-l}^{u} dp_\tau dp_u\right)\\ &= 4E\left(\int_{t-s}^{t}\int_{t-s}^{u} dp_\tau dp_u E\left[\int_{t+\delta-l}^{t+\delta}\int_{t+\delta-l}^{u} dp_\tau dp_u \middle| p_{z,0\le z\le t}\right]\right)\\ &= 0.\end{aligned} \tag{A.11}$$

For $\delta < l$, $l - \delta < s$, using (A.10): $Z_{t,s} = Z_{t,l-\delta} + Z_{t-l+\delta,s+\delta-l} + 2r^*{}_{t,l-\delta}r^*{}_{t-l+\delta,s+\delta-l}$, $Z_{t+\delta,l} = Z_{t+\delta,\delta} + Z_{t,l-\delta} + 2r^*{}_{t+\delta,\delta}r^*{}_{t,l-\delta}$. Using (A.11), $\mathrm{Cov}(Z_{t,s}, Z_{t+\delta,l}) = \mathrm{Var}(Z_{t,l-\delta})$. For $\delta < l$, $l - \delta \ge s$, by analogy, $\mathrm{Cov}(Z_{t,s}, Z_{t+\delta,l}) = \mathrm{Var}(Z_{t,s})$. □

Lemma A.2. *Define the finest (every second) time grid $\mathcal{M} = \{t_0, t_1, \ldots, t_M\}$ and m subgrids $\mathcal{M}_i^m = \{t'_0, t'_1, t'_2, \ldots, t'_{\overline{M}}\}$, $t'_k \equiv t_{i+km}$, $i \in \{0, 1, 2, \ldots, m-1\}$, with corresponding $r_{t_k,m} \equiv p_{t_k} - p_{t_{k-m}}$ and $e_{t_k,m} \equiv \eta_{t_k} - \eta_{t_{k-m}}$. Further, define $h \equiv m/M$ Then, $\forall \mathcal{M}_i^m, \mathcal{M}_j^m, i \neq j, \mathcal{M}_j^m \bigcap \mathcal{M}_i^m = \emptyset$ and:*

$$\mathrm{Var}\left(\frac{1}{m}\sum_{i=0}^{m-1}\sum_{t_{k,-1}\in\mathcal{M}_i^m} e_{t_k}^2\right) = \frac{1}{m}\mathrm{Var}\left(\sum_{t_{j,-1}\in\mathcal{M}_i^m} e_{t_k}^2\right); \qquad \text{(A.12)}$$

$$\mathrm{Var}\left(\frac{1}{m}\sum_{i=0}^{m-1}\sum_{t_{k,-1}\in\mathcal{M}_i^m} e_{t_k} r_{t_k}^*\right) = \frac{1}{m}\mathrm{Var}\left(\sum_{t_{k,-1}\in\mathcal{M}_i^m} e_{t_k} r_{t_k}^*\right)$$

$$\mathrm{Cov}\left(\sum_{t_{k,-1}\in\mathcal{M}} e_{t_k}^2, \sum_{t_{k,-1}\in\mathcal{M}_j^m} e_{t_k}^2\right) \equiv \mathrm{Cov}\left(\sum_{i=1}^{M} e^2_{\frac{i}{M},\frac{1}{M}}, \sum_{i=1}^{\overline{M}} e^2_{\frac{j}{M}+ih,h}\right)$$

$$= \begin{cases}(4\overline{M}-1)\mathrm{Var}\left(\eta_0^2\right) & j = \{0, m-1\}\\ 4\overline{M}\mathrm{Var}\left(\eta_0^2\right) & j = \{1, \ldots, m-2\}\end{cases} \qquad \text{(A.13)}$$

$$\mathrm{Var}\left(\sum_{t_{k,-1}\in\mathcal{M}_j^m} e_{t_k}^2\right) \equiv \mathrm{Var}\left(\sum_{i=1}^{\overline{M}} e^2_{\frac{j}{M}+ih,h}\right)$$

$$= (4\overline{M}-2)\mathrm{Var}\left(\eta_0^2\right) + 4\overline{M}\sigma_\eta^4, \qquad \text{(A.14)}$$

$$\mathrm{Var}\left(\sum_{t_{k,-1}\in\mathcal{M}_i^m} e_{t_k} r_{t_k}^*\right) = 2a_0 h\overline{M}\sigma_\eta^2$$

$$\textit{if } i \geq j, \quad \mathrm{Cov}\left(\sum_{k=1}^{\overline{M}} Z_{\frac{i}{M}+kh,h}, \sum_{k=1}^{\overline{M}} Z_{\frac{j}{M}+kh,h}\right)$$

$$= (\overline{M}-1)\mathrm{Var}\left(Z_{\frac{i}{M}+h,\frac{i-j}{M}}\right) + \overline{M}\mathrm{Var}\left(Z_{\frac{j}{M}+h,h-\frac{i-j}{M}}\right) \qquad \text{(A.15)}$$

$$\mathrm{Cov}\left(\sum_{k=1}^{\overline{M}} Z_{\frac{i}{M}+kh,h}, \sum_{k=1}^{M} Z_{k/M,\frac{1}{M}}\right) = (M-m)\mathrm{Var}\left(Z_{i,\frac{1}{M}}\right). \qquad \text{(A.16)}$$

Proof. Eqs. (A.12) hold since $\forall \mathcal{M}_i^m, \mathcal{M}_j^m, i \neq j, \mathcal{M}_j^m \bigcap \mathcal{M}_i^m = \emptyset$; η_t is *i.i.d.*; η_t and r_t^* are independent

$$\mathrm{E}\left(\sum_{t_{k,-1}\in\mathcal{M}_i^m} e_{t_k}^2 \sum_{t_{k,-1}\in\mathcal{M}_j^m} e_{t_k}^2\right) = \mathrm{E}\left(\sum_{t_{k,-1}\in\mathcal{M}_i^m} e_{t_k}^2\right)\mathrm{E}\left(\sum_{t_{k,-1}\in\mathcal{M}_j^m} e_{t_k}^2\right)$$

$$\mathrm{E}\left(\sum_{t_{k,-1}\in\mathcal{M}_i^m} e_{t_k} r_{t_k}^* \sum_{t_{k,-1}\in\mathcal{M}_j^m} e_{t_k} r_{t_k}^*\right)$$

$$= \mathrm{E}\left(\sum_{t_{k,-1}\in\mathcal{M}_i^m} e_{t_k} r_{t_k}^* \sum_{t_{k,-1}\in\mathcal{M}_j^m} \mathrm{E}\left(e_{t_k}|\eta_{t_k}, t_k \in \mathcal{M}_i^m\right) r_{t_k}^*\right) = 0$$

(A.13): $\mathrm{Cov}\left(\eta_0^2 + \eta_1^2 + 2\sum_{i=1}^{M-1}\eta^2_{\frac{i}{M}}, \eta^2_{\frac{j}{M}}\right.$

$$\left. + \ \eta^2_{1-(m-j-1)/M} + 2\sum_{i=1}^{\overline{M}-1}\eta^2_{\frac{j}{M}+ih}\right)$$

$$= 4(\overline{M}-1)\mathrm{Var}\left(\eta_0^2\right) + \mathrm{Cov}\left(\eta_0^2 + \eta_1^2 + 2\sum_{i=1}^{M-1}\eta^2_{\frac{i}{M}},\right.$$

$$\left.\eta^2_{\frac{j}{M}} + \eta^2_{1-(j-m+1)/M}\right)$$

$$= \begin{cases}(4\overline{M}-1)\mathrm{Var}\left(\eta_0^2\right) & j = \{0, m-1\}\\ 4\overline{M}\mathrm{Var}\left(\eta_0^2\right) & j = \{1, \ldots, m-2\}\end{cases}$$

(A.14): $\mathrm{Var}\left(\eta^2_{\frac{j}{M}} + \eta^2_{1-(m-j-1)/M} + 2\sum_{i=1}^{\overline{M}-1}\eta^2_{\frac{j}{M}+ih}\right.$

$$\left. - \ 2\sum_{j=1}^{\overline{M}}\eta_{\frac{j}{M}+ih}\eta_{\frac{j}{M}+(i-1)h}\right)$$

$$= (4\overline{M}-2)\mathrm{Var}\left(\eta_0^2\right) + 4\overline{M}\sigma_\eta^4$$

$$\mathrm{Var}\left(\sum_{i=1}^{\overline{M}} e_{\frac{j}{M}+ih,h} r^*_{\frac{j}{M}+ih,h}\right)$$

$$= \mathrm{E}\left(\sum_{i=1}^{\overline{M}}\left(\eta^2_{\frac{j}{M}+ih} - 2\eta_{\frac{j}{M}+ih}\eta_{\frac{j}{M}+(i-1)h} + \eta^2_{\frac{j}{M}+(i-1)h}\right) r^{*2}_{\frac{j}{M}+ih,h}\right)$$

$$= 2a_0 h\overline{M}\sigma_\eta^2$$

(A.15) using Lemma A.1: $\mathrm{Cov}\left(Z_{\frac{j}{M}+h,h}, \sum_{k=1}^{\overline{M}} Z_{\frac{i}{M}+kh,h}\right)$

$$= \mathrm{Cov}\left(Z_{\frac{j}{M}+h,h}, Z_{\frac{i}{M}+h,h}\right)$$

$$= \mathrm{Var}\left(Z_{\frac{j}{M}+h,h-\frac{i-j}{M}}\right)$$

$$\forall k > 1, \quad \mathrm{Cov}\left(Z_{\frac{j}{M}+kh,h}, \sum_{k=1}^{\overline{M}} Z_{\frac{i}{M}+kh,h}\right)$$

$$= \mathrm{Cov}\left(Z_{\frac{j}{M}+kh,h}, Z_{\frac{i}{M}+kh,h}\right) + \mathrm{Cov}\left(Z_{\frac{j}{M}+kh,h}, Z_{\frac{i}{M}+(k-1)h,h}\right)$$

$$= \mathrm{Var}\left(Z_{\frac{j}{M}+h,h-\frac{i-j}{M}}\right) + \mathrm{Var}\left(Z_{\frac{j}{M}+h,\frac{i-j}{M}}\right)$$

$$\sum_{k=1}^{\overline{M}}\mathrm{Cov}\left(Z_{\frac{j}{M}+kh,h}, \sum_{k=1}^{\overline{M}} Z_{\frac{i}{M}+kh,h}\right)$$

$$= (\overline{M}-1)\mathrm{Var}\left(Z_{\frac{i}{M}+h,\frac{i-j}{M}}\right) + \overline{M}\mathrm{Var}\left(Z_{\frac{j}{M}+h,h-\frac{i-j}{M}}\right).$$

Similarly, Eq. (A.16): $\mathrm{Cov}\left(Z_{\frac{l}{M},\frac{1}{M}}, \sum_{k=1}^{\overline{M}} Z_{\frac{i}{M}+kh,h}\right)$

$$= \begin{cases}0, & l \in [i, M-m+i+1)\\ \mathrm{Var}\left(Z_{\frac{1}{M},\frac{1}{M}}\right) & \text{otherwise}\end{cases}$$

$$\sum_{l=1}^{M}\mathrm{Cov}\left(Z_{\frac{l}{M},\frac{1}{M}}, \sum_{k=1}^{\overline{M}} Z_{\frac{i}{M}+kh,h}\right) = (M-m)\mathrm{Var}\left(Z_{\frac{1}{M},\frac{1}{M}}\right). \quad \square$$

Lemma A.3. *Given conditions above,*

$$\mathrm{Var}\left(\frac{1}{m}\sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}} r^{*2}_{\frac{j}{M}+kh,h}\right) = \mathrm{Var}\left(\frac{1}{m}\sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}} IV^{*2}_{\frac{j}{M}+kh,h}\right) + \mathrm{Var}\left(\frac{1}{m}\sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}} Z^{*2}_{\frac{j}{M}+kh,h}\right). \qquad \text{(A.17)}$$

Proof of Theorem 2.1. Note that by construction, $M+1$, the number of observations in the finest grid $\mathcal{M}$ equal to $\overline{M}m+m$, where $\overline{M}+1$ is the number of observations in the grid $\mathcal{M}_j^m$. As a result, $\overline{M}m = M+1-m$, or, given $h = m/M$, $\overline{M}h = (M+1-m)/M$. Sum of efficient squared returns over some period can be separated into two parts: *IV* part and discretization error *Z* part.

$$\sum_{k=1}^{\overline{M}} r^{*2}_{\frac{j}{M}+kh,h} = \sum_{k=1}^{\overline{M}} IV_{\frac{j}{M}+kh,h} + \sum_{k=1}^{\overline{M}} Z_{\frac{j}{M}+kh,h} = IV_{\frac{j}{M}+(M-m+1)/M,(M-m+1)/M} + \sum_{k=1}^{\overline{M}} Z_{\frac{j}{M}+kh,h}. \quad (A.18)$$

Then, using results of Lemmas A.1 and A.2, properties of integrated variance and discretization noise (A.9)

$$\begin{aligned}
\operatorname{Var}(\mathrm{RV}_j^m) &= \operatorname{Var}\left(\sum_{k=1}^{\overline{M}} r^{*2}_{\frac{j}{M}+kh,h}\right) + 4\operatorname{Var}\left(\sum_{k=1}^{\overline{M}} e_{\frac{j}{M}+kh,h} r^{*}_{\frac{j}{M}+kh,h}\right) + \operatorname{Var}\left(\sum_{k=1}^{\overline{M}} e^2_{\frac{j}{M}+kh,h}\right)\\
&= \operatorname{Var}\left(IV_{\overline{M}h+\frac{j}{M},\overline{M}h}\right) + \overline{M}\operatorname{Var}\left(Z_{\frac{j}{M}+h,h}\right) + 8\overline{M}a_0 h\sigma_\eta^2 + (4\overline{M}-2)\operatorname{Var}\left(\eta_0^2\right) + 4\overline{M}\sigma_\eta^4\\
&= 2\sum_{i=1}^{p}\frac{a_i^2}{\lambda_i^2}\left(e^{-\lambda_i \overline{M}h} - 1 + \lambda_i \overline{M}h\right) + 4\overline{M} \times \left[\frac{a_0^2h^2}{2} + \sum_{i=1}^{p}\frac{a_i^2}{\lambda_i^2}\left(e^{-\lambda_i h} - 1 + \lambda_i h\right) + 2a_0h\sigma_\eta^2 + \sigma_\eta^4\right] + (4\overline{M}-2)(\kappa-1)\sigma_\eta^4 \quad (A.19)
\end{aligned}$$

$$\operatorname{Var}(\mathrm{RV}^m_{AC_1}) = \operatorname{Var}(\hat\gamma_0) + 4\frac{\overline{M}^2}{(\overline{M}-1)^2}\operatorname{Var}(\hat\gamma_1) + 4\frac{\overline{M}}{(\overline{M}-1)}\operatorname{Cov}(\hat\gamma_0,\hat\gamma_1), \quad (A.20)$$

where $\operatorname{Var}(\hat\gamma_0)$ is (A.19), $\operatorname{Var}(\hat\gamma_1)$ and $\operatorname{Cov}(\hat\gamma_0,\hat\gamma_1)$ computed using appropriate modification of BHLS appendix (p. 25), i.e. $\operatorname{Cov}(\hat\gamma_0,\hat\gamma_1) = -2(\overline{M}-1)[(\kappa+1)\sigma_\eta^4 + 2a_0h\sigma_\eta^2]$, $\operatorname{Var}(\hat\gamma_1) = (\overline{M}-1)[(\kappa+2)\sigma_\eta^4 + 4a_0h\sigma_\eta^2 + a_0^2h^2 + \sum_{i=1}^{p}\frac{a_i^2}{\lambda_i^2}(1-e^{-\lambda_i h})^2] + 2(\overline{M}-2)\sigma_\eta^4$. Variance of the averaging over subsamples estimator (A.5) is

$$\operatorname{Var}\left(\overline{\mathrm{RV}}^m\right) = \operatorname{Var}\left(\frac{1}{m}\sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}} r^{*2}_{\frac{j}{M}+kh,h}\right) + 4\operatorname{Var}\left(\frac{1}{m}\sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}} e_{\frac{j}{M}+kh,h} r^{*}_{\frac{j}{M}+kh,h}\right) + \operatorname{Var}\left(\frac{1}{m}\sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}} e^2_{\frac{j}{M}+kh,h}\right).$$

Using Lemma A.2, $\operatorname{Var}\left(\overline{\mathrm{RV}}^m\right) = \operatorname{Var}\left(\frac{1}{m}\sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}} r^{*2}_{\frac{j}{M}+kh,h}\right) + \frac{8}{m}a_0h\overline{M}\sigma_\eta^2 + \frac{1}{m}((4\overline{M}-2)\operatorname{Var}\left(\eta_0^2\right) + 4\overline{M}\sigma_\eta^4)$

$$\operatorname{Var}\left(\frac{1}{m}\sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}} r^{*2}_{\frac{j}{M}+kh,h}\right) = \operatorname{Var}\left(\frac{1}{m}\sum_{j=0}^{m-1} IV_{1-\frac{i}{M},1-\frac{2i}{M}}\right) + \operatorname{Var}\left(\frac{1}{m}\sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}} Z_{\frac{j}{M}+kh,h}\right), \quad \text{with}$$

$$\begin{aligned}
\operatorname{Var}\left(\frac{1}{m}\sum_{j=0}^{m-1} IV_{\frac{M-i}{M},\frac{M-2i}{M}}\right) &= \frac{1}{m^2}\sum_{i=0}^{m-1}(1+2i)\operatorname{Var}\left(IV_{1-\frac{i}{M},1-\frac{2i}{M}}\right) + \frac{4}{m^2}\sum_{i=1}^{m-1}\sum_{j=0}^{i-1}\operatorname{Cov}\left(IV_{1-\frac{i}{M},1-\frac{2i}{M}}, IV_{1-\frac{j}{M},\frac{i-j}{M}}\right)\\
&= \frac{1}{m^2}\left(\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\sum_{i=0}^{m-1}(1+2i) \times \left[e^{-\lambda_k(1-\frac{2i}{M})} - 1 + \lambda_k\left(1-\frac{2i}{M}\right)\right]\right) + \frac{4}{m^2}\sum_{i=1}^{m-1}\sum_{j=0}^{i-1} \times \left(\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left[1-e^{-\lambda_k(1-\frac{2i}{M})}\right]\left[1-e^{-\lambda_k\frac{i-j}{M}}\right]\right), \quad \text{and}
\end{aligned}$$

$$\begin{aligned}
\operatorname{Var}\left(\frac{1}{m}\sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}} Z_{\frac{j}{M}+kh,h}\right) &= \frac{1}{m}\overline{M}\operatorname{Var}\left(Z_{h,h}\right) + \frac{2}{m^2}\sum_{i=1}^{m-1}\sum_{j=0}^{i-1}\operatorname{Cov}\left(\sum_{k=1}^{\overline{M}} Z_{\frac{j}{M}+kh,h}, \sum_{k=1}^{\overline{M}} Z_{\frac{i}{M}+kh,h}\right)\\
&= \frac{2}{m^2}\sum_{i=1}^{m-1}\sum_{j=0}^{i-1}\left((\overline{M}-1)\operatorname{Var}\left(Z_{\frac{i}{M}+h,\frac{i-j}{M}}\right) + \overline{M}\operatorname{Var}\left(Z_{\frac{j}{M}+h,h-\frac{i-j}{M}}\right)\right) + \frac{\overline{M}}{m}\operatorname{Var}\left(Z_{h,h}\right)\\
&= \frac{2}{m^2}\sum_{k=1}^{m-1}\left((m-k)(\overline{M}-1)\operatorname{Var}\left(Z_{h,k/M}\right) + \overline{M}k\operatorname{Var}\left(Z_{h,k/M}\right)\right) + \frac{\overline{M}}{m}\operatorname{Var}\left(Z_{h,h}\right)\\
&= \frac{2(\overline{M}-1)}{m}\sum_{i=1}^{m-1}\operatorname{Var}\left(Z_{h,\frac{i}{M}}\right) + \frac{2}{m^2} \times \sum_{i=1}^{m-1} i\operatorname{Var}\left(Z_{h,\frac{i}{M}}\right) + \frac{\overline{M}}{m}\operatorname{Var}\left(Z_{h,h}\right)\\
&= \frac{8(\overline{M}-1)}{m}\sum_{i=1}^{m-1}\left\{\frac{i^2a_0^2}{2M^2} + \sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(e^{-\lambda_k\frac{i}{M}} - 1 + \lambda_k\frac{i}{M}\right)\right\} + \frac{8}{m^2}\sum_{i=1}^{m-1} i\left\{\frac{i^2a_0^2}{2M^2} + \sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(e^{-\lambda_k\frac{i}{M}} - 1 + \lambda_k\frac{i}{M}\right)\right\}\\
&\quad + \frac{\overline{M}}{m}\left(2h^2a_0^2 + 4\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(e^{-\lambda_k h} - 1 + \lambda_k h\right)\right). \quad (A.21)
\end{aligned}$$

With the final result

$$\begin{aligned}
\operatorname{Var}\left(\overline{\mathrm{RV}}^m\right) &= \frac{8a_0h\overline{M}\sigma_\eta^2}{m} + \frac{(4\overline{M}-2)(\kappa-1)\sigma_\eta^4 + 4\overline{M}\sigma_\eta^4}{m} + \frac{1}{m^2}\left(\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\sum_{i=0}^{m-1}(1+2i)\left[e^{-\lambda_k\frac{M-2i}{M}} - 1 + \lambda_k\frac{M-2i}{M}\right]\right)\\
&\quad + \frac{4}{m^2}\sum_{i=1}^{m-1}\sum_{j=0}^{i-1}\left(\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left[1-e^{-\lambda_k\frac{M-2i}{M}}\right]\left[1-e^{-\lambda_k\frac{i-j}{M}}\right]\right) + \frac{8(\overline{M}-1)}{m}\sum_{i=1}^{m-1}\left\{\frac{i^2a_0^2}{2M^2} + \sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(e^{-\lambda_k\frac{i-j}{M}} - 1 + \frac{\lambda_k i}{M}\right)\right\}\\
&\quad + \frac{8}{m^2}\sum_{i=1}^{m-1} i\left\{\frac{i^2a_0^2}{2M^2} + \sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(e^{-\lambda_k\frac{i}{M}} - 1 + \lambda_k\frac{i}{M}\right)\right\}
\end{aligned}$$

$$+\frac{\overline{M}}{m}\left(2h^2a_0^2+4\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(\mathrm{e}^{-\lambda_k h}-1+\lambda_k h\right)\right). \quad (A.22)$$

And finally, variance of the two-scales estimator (A.4) with small-sample correction is

$$\left(1-\frac{\overline{M}}{M}\right)^2\mathrm{Var}\left(\mathrm{RV}_{TS}\right)=\mathrm{Var}\left(\overline{\mathrm{RV}}^m\right)+\frac{\overline{M}^2}{M^2}\mathrm{Var}\left(\mathrm{RV}^1\right)-2\frac{\overline{M}}{M}\mathrm{Cov}\left(\overline{\mathrm{RV}}^m,\mathrm{RV}^1\right). \quad (A.23)$$

The first and the second terms of the variance is already computed.

$$\begin{aligned}
m\mathrm{Cov}\left(\overline{\mathrm{RV}}^m,\mathrm{RV}^1\right)&=\mathrm{Cov}\Bigg(\sum_{i=1}^{M}r^{*2}_{\frac{i}{M},\frac{1}{M}}+2r^*_{\frac{i}{M},\frac{1}{M}}e_{\frac{i}{M},\frac{1}{M}}+e^2_{\frac{i}{M},\frac{1}{M}},\\
&\quad\sum_{j=0}^{m-1}\left\{\sum_{k=1}^{\overline{M}}r^{*2}_{\frac{j}{M}+kh,h}+2\sum_{k=1}^{\overline{M}}e_{\frac{j}{M}+kh,h}r^*_{\frac{j}{M}+kh,h}+e^2_{\frac{j}{M}+kh,h}\right\}\Bigg)\\
&\quad\times\mathrm{Cov}\left(\sum_{i=1}^{M}r^{*2}_{\frac{i}{M},\frac{1}{M}},\sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}}r^{*2}_{\frac{j}{M}+kh,h}\right)\\
&\quad+4\mathrm{Cov}\left(\sum_{i=1}^{M}r^*_{\frac{i}{M},\frac{1}{M}}e_{\frac{i}{M},\frac{1}{M}},\sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}}r^*_{\frac{j}{M}+kh,h}e_{\frac{j}{M}+kh,h}\right)\\
&\quad+\mathrm{Cov}\left(\sum_{i=1}^{M}e^2_{\frac{i}{M},\frac{1}{M}},\sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}}e^2_{\frac{j}{M}+kh,h}\right)\\
&\equiv mA+mB+mC.
\end{aligned}$$

Using Lemma A.2,

$$\begin{aligned}
C&=\frac{1}{m}\sum_{j=0}^{m-1}\mathrm{Cov}\left(\sum_{i=1}^{M}e^2_{\frac{i}{M},\frac{1}{M}},\sum_{k=1}^{\overline{M}}e^2_{\frac{j}{M}+kh,h}\right)\\
&=\left(4\overline{M}-2/m\right)(\kappa-1)\sigma^4_\eta\\
B&=\frac{4}{m}\sum_{j=0}^{m-1}\mathrm{Cov}\left(\sum_{i=1}^{M}r^*_{\frac{i}{M},\frac{1}{M}}e_{\frac{i}{M},\frac{1}{M}},\sum_{k=1}^{\overline{M}}r^*_{\frac{j}{M}+kh,h}e_{\frac{j}{M}+kh,h}\right)\\
&=\frac{4}{m}\sum_{j=0}^{m-1}\mathrm{Cov}\left(\sum_{i=1}^{M}r^*_{\frac{i}{M},\frac{1}{M}}\eta_{\frac{i}{M}},\sum_{k=1}^{\overline{M}}r^*_{\frac{j}{M}+kh,h}\eta_{\frac{j}{M}+kh}\right)\\
&=\frac{4a_0\overline{M}\sigma^2_\eta}{M}\\
A&=\frac{1}{m}\sum_{j=0}^{m-1}\mathrm{Cov}\left(\sum_{i=1}^{M}r^{*2}_{\frac{i}{M},\frac{1}{M}},\sum_{k=1}^{\overline{M}}r^{*2}_{\frac{j}{M}+kh,h}\right)\\
&=\frac{1}{m}\sum_{j=0}^{m-1}\mathrm{Cov}\left(IV_{1,1},IV_{\overline{M}h+\frac{j}{M},\overline{M}h}\right)\\
&\quad+\frac{1}{m}\sum_{j=0}^{m-1}\mathrm{Cov}\left(\sum_{i=1}^{M}Z_{\frac{i}{M},\frac{1}{M}},\sum_{k=1}^{\overline{M}}Z_{\frac{j}{M}+kh,h}\right)\\
&=\frac{1}{m}\sum_{j=0}^{m-1}\mathrm{Cov}\Big(IV_{\frac{j}{M},\frac{j}{M}}+IV_{\overline{M}h+\frac{j}{M},\overline{M}h}\\
&\quad+IV_{1,(m-1-j)/M},IV_{\overline{M}h+\frac{j}{M},\overline{M}h}\Big)+(M-m)\mathrm{Var}\left(Z_{\frac{1}{M},\frac{1}{M}}\right)\\
&=\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(\left(1-\frac{2(1-\mathrm{e}^{-\lambda_k h})}{m\left(1-\mathrm{e}^{-\lambda_k/M}\right)}\right)(1-\mathrm{e}^{-\lambda_k\overline{M}h})+\lambda_k\overline{M}h\right)\\
&\quad+(M-m)\left(\frac{2a_0^2}{M^2}+4\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}(\mathrm{e}^{-\lambda_k/M}-1+\lambda_k/M)\right).\quad\square
\end{aligned}$$

Proof of Theorem 2.2. Given $\mathrm{Cov}\left(\widehat{\mathrm{RV}}^i_{t+a,a},\widehat{\mathrm{RV}}^j_{t-s,b}\right)=\mathrm{Cov}\left(IV_{t+a,a},IV_{t-s,b}\right)$, $s\geq 0$, the vector $C\left(\widehat{\mathrm{RV}}^i_{t+nh,nh},\widehat{\mathrm{RV}}^j_{t,h},l\right)$ can be rewritten as $C\left(\widehat{\mathrm{RV}}^i_{t+nh,nh},\widehat{\mathrm{RV}}^j_{t,h},l\right)=C\left(IV_{t+nh,nh},IV_{t,h},l\right)\equiv C(.)$, and the matrix $M(\widehat{\mathrm{RV}}^i_{t,h},l)$ as $M(\widehat{\mathrm{RV}}^i_{t,h},l)=M(IV_{t,h},l)+I\left[\mathrm{Var}\left(\widehat{\mathrm{RV}}^i_{t,h}\right)-\mathrm{Var}\left(IV_{t,h}\right)\right]$ where I is $(l+1)\times(l+1)$ identity matrix. Without loss of generality we can assume that

$$\begin{aligned}
&\mathrm{Var}\left(\widehat{\mathrm{RV}}^X_{t,h}\right)=\mathrm{Var}\left(\widehat{\mathrm{RV}}^Y_{t,h}\right)+\delta,\ \delta\geq 0\Leftrightarrow M\left(\widehat{\mathrm{RV}}^X_{t,h}\right)\\
&\quad=M\left(\widehat{\mathrm{RV}}^Y_{t,h}\right)+I\delta\Rightarrow R^2(\widehat{\mathrm{RV}}^Z_{t+nh,nh},\widehat{\mathrm{RV}}^Y_{t,h},l)\\
&\quad\geq R^2(\widehat{\mathrm{RV}}^Z_{t+nh,nh},\widehat{\mathrm{RV}}^X_{t,h},l)\\
&\Leftrightarrow C(.)'\left[M\left(\widehat{\mathrm{RV}}^Y_{t,h}\right)\right]^{-1}C(.)-C(.)'\left[M\left(\widehat{\mathrm{RV}}^X_{t,h}\right)\right]^{-1}C(.)\geq 0\\
&\Leftrightarrow C(.)'\left\{\left[M\left(\widehat{\mathrm{RV}}^Y_{t,h}\right)\right]^{-1}-\left[M\left(\widehat{\mathrm{RV}}^X_{t,h}\right)\right]^{-1}\right\}C(.)\geq 0\\
&\Leftrightarrow\left[M\left(\widehat{\mathrm{RV}}^Y_{t,h}\right)\right]^{-1}-\left[M\left(\widehat{\mathrm{RV}}^Y_{t,h}\right)+I\delta\right]^{-1}\ \text{p.s.d.}\\
&\Leftrightarrow\left[M\left(\widehat{\mathrm{RV}}^Y_{t,h}\right)+I\delta\right]\left[M\left(\widehat{\mathrm{RV}}^Y_{t,h}\right)\right]^{-1}\\
&\qquad\times\left[M\left(\widehat{\mathrm{RV}}^Y_{t,h}\right)+I\delta\right]-M\left(\widehat{\mathrm{RV}}^Y_{t,h}\right)-I\delta\\
&\quad=\delta^2\left[M\left(\widehat{\mathrm{RV}}^Y_{t,h}\right)\right]^{-1}+I\delta\quad\text{p.s.d. by construction.}\quad\square
\end{aligned}$$

For numerical computations we are using models M1–M3 described in Andersen et al. (2004).

Model M1—GARCH diffusion, popularized by Nelson (1990): $\mathrm{d}\sigma_t^2=\kappa(\theta-\sigma_t^2)\mathrm{d}t+\sigma\sigma_t^2\mathrm{d}W_t$.

The spot volatility of this process can be expressed as

$$\sigma_t^2=\theta+\theta\sqrt{\frac{\sigma^2}{2\kappa-\sigma^2}}P_1(f_t) \quad (A.24)$$

with $\lambda_1=\kappa$, $P_1(f_t)=\frac{\sqrt{2\kappa-\sigma^2}}{\theta\sqrt{\sigma^2}}(f_t-\theta)$, and $\mathrm{d}f_t=\kappa(\theta-f_t)\mathrm{d}t+\sigma f_t\mathrm{d}W_t$. For simulations we use $a_0=0.686$, $a_1=0.412$, $\lambda_1=0.035$.

Model M2—Two-factor affine: $\sigma_t^2=\sigma_{1,t}^2+\sigma_{2,t}^2$, $\mathrm{d}\sigma_{j,t}^2=\kappa_j(\theta_j-\sigma_{j,t}^2)\mathrm{d}t+\eta_j\sigma_{j,t}\mathrm{d}W_t^j$, $j=1,2$.

The spot variance of this process can be expressed as

$$\sigma_t^2=(\theta_1+\theta_2)-\frac{\theta_1}{\sqrt{\alpha_1}}L_1^{(\alpha_1-1)}(f_{1,t})-\frac{\theta_2}{\sqrt{\alpha_2}}L_2^{(\alpha_2-1)}(f_{2,t}) \quad (A.25)$$

with $L_1^{(\alpha_j-1)}(f_{j,t})$ are the Laguerre polynomials of degree 1 with corresponding eigenvalues $\lambda_j=\kappa_j, f_{j,t}=\alpha_j/\theta_j\sigma_{j,t}^2$, $\mathrm{d}f_{j,t}=\kappa_j(\alpha_j-f_{j,t})\mathrm{d}t+\sqrt{2\kappa_jf_{j,t}}\mathrm{d}W_t^j$, and $\alpha_j=2\kappa_j\theta_j/\eta_j^2$. For simulations we use $a_0=0.504$, $a_1=-0.122$, $a_2=-0.119$, $\lambda_1=0.571$, $\lambda_2=0.076$.

Model M3—Log-normal diffusion: $\mathrm{d}\log(\sigma_t^2)=\kappa[\theta-\log(\sigma_t^2)]\mathrm{d}t+\sigma\mathrm{d}W_t$.

The spot variance of this process can be expressed as

$$\sigma_t^2=\sum_{i=0}^{\infty}a_iH_i(f_t) \quad (A.26)$$

where $H_i(f_t), i=0,1,\ldots$ are Hermite polynomial with corresponding eigenvalues $\lambda_i=\kappa i$, $a_i=\exp(\theta+\sigma^2/4\kappa)(\sigma/\sqrt{2\kappa})^i/\sqrt{i!}$, and $f_t=\sqrt{2\kappa}(\log\sigma_t^2-\theta)/\sigma$. For simulations we use $a_0=0.551$, $a_1=0.387$, $a_n=\frac{a_1^n}{a_0^{n-1}\sqrt{n!}}$, $\lambda_1=0.014$, $\lambda_n=\lambda_1 n$.

A.3. Proof of Proposition 2.1

Defining $Q = \sum_{i=0}^{p} a_i^2$, $\phi = \overline{M}/M$ and assuming $\overline{M}h \sim 1$, $m \sim M/\overline{M}$, $\mathrm{Var}\left(Z_{t,h}\right) \sim 2h^2 Q$, variances of realized volatility estimators RV, RV_{AC1}, RV_{TS}, $\overline{\mathrm{RV}}$ as a function of $\overline{M}$ can be approximated by:

$$
\begin{aligned}
\mathrm{Var}\left(\mathrm{RV}_j^{(m)}\right) &\simeq \mathrm{Var}\left(IV_{1,1}\right) + \frac{2Q}{\overline{M}} + 8a_0\sigma_\eta^2 \\
&\quad + 4\overline{M}\kappa\sigma_\eta^4 - 2\mathrm{Var}\left(\eta^2\right) \\
\mathrm{Var}\left(\mathrm{RV}_{AC_1}\right) &\simeq C + \frac{2Q}{\overline{M}} + 4\overline{M}\kappa\sigma_\eta^4 + 4\overline{M}\left[(\kappa+2)\sigma_\eta^4 + h^2 Q\right] \\
&\quad + 8\overline{M}\sigma_\eta^4 - 8\overline{M}(\kappa+1)\sigma_\eta^4 = C + \frac{6Q}{\overline{M}} + 8\overline{M}\sigma_\eta^4 \\
\mathrm{Var}\left(\overline{\mathrm{RV}}\right) &\simeq \mathrm{Var}\left(IV_{1,1}\right) + 8a_0\phi\sigma_\eta^2 \\
&\quad + 4\phi^2 M\kappa\sigma_\eta^4 + \frac{2Q}{M} - 2\phi\mathrm{Var}\left(\eta^2\right) \\
&\quad + \frac{Q(1-\phi)\left\{2M(2-\phi) - \frac{1+\phi}{\phi}\right\}}{3M^2\phi} \\
&= \mathrm{Var}\left(IV_{1,1}\right) + 8a_0\phi\sigma_\eta^2 + 4\phi^2 M\kappa\sigma_\eta^4 \\
&\quad - 2\phi\mathrm{Var}\left(\eta^2\right) + \frac{2Q}{3M}\left(\frac{2}{\phi} + \phi\right) - \frac{Q}{3M^2}\left(\frac{1}{\phi^2} - 1\right)
\end{aligned}
$$

$$
\begin{aligned}
(1-\phi)^2\,\mathrm{Var}\,(\mathrm{RV}_{TS}) &\simeq (1-\phi)^2\left(\mathrm{Var}\left(IV_{1,1}\right) + \frac{2Q}{M}\right) \\
&\quad + \frac{Q(1-\phi)\left\{2M(2-\phi) - \frac{1+\phi}{\phi}\right\}}{3M^2\phi} \\
&\quad + 8\phi a_0\sigma_\eta^2 + 8\phi^2 M\sigma_\eta^4 - 2\phi(1-\phi)\mathrm{Var}\left(\eta^2\right) + \frac{4Q}{M^2},
\end{aligned}
$$

$$
\begin{aligned}
\text{or}\quad \mathrm{Var}\,(\mathrm{RV}_{TS}) &\simeq \left(\mathrm{Var}\left(IV_{1,1}\right) + \frac{2Q}{M}\right) \\
&\quad + \frac{Q\left\{2M(2-\phi) - (1/\phi + 1)\right\}}{3M^2\phi(1-\phi)} - \frac{2\phi\mathrm{Var}\left(\eta^2\right)}{1-\phi} \\
&\quad + \frac{1}{(1-\phi)^2}\left(\frac{4Q}{M^2} + 8\phi a_0\sigma_\eta^2 + 8\phi^2 M\sigma_\eta^4\right). \quad \square \qquad \text{(A.27)}
\end{aligned}
$$

References

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

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

Aït-Sahalia, Yacine, Yu, Jialin, 2009. High frequency market microstructure noise estimates and liquidity measures. Annals of Applied Statistics 3, 422–457.

Andersen, Torben, Bollerslev, Tim, Meddahi, Nour, 2004. Analytic evaluation of volatility forecasts. International Economic Review 45, 1079–1110.

Andersen, Torben, Bollerslev, Tim, Meddahi, Nour, 2011. Realized volatility forecasting and market microstructure noise. Journal of Econometrics 160 (1), 220–234.

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

Bandi, Federico, Russell, Jeffrey, 2008a. Microstructure noise, realized variance, and optimal sampling. Review of Economic Studies 75, 339–369.

Bandi, Frederico M., Russell, Jeffrey R., 2008b. On the finite sample properties of kernel-based integrated variance estimators. Working Paper.

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

Garcia, Rene, Meddahi, Nour, 2006. Comment. Journal of Business and Economic Statistics 24 (2), 184–192.

Ghysels, Eric, Santa-Clara, Pedro, Valkanov, Rossen, 2006. Predicting volatility: getting the most out of return data sampled at different frequencies. Journal of Econometrics 131, 59–95.

Ghysels, Eric, Sinko, Arthur, 2006. Comment. Journal of Business and Economic Statistics 24 (2), 192–194.

Ghysels, Eric, Sinko, Arthur, 2009. Volatility forecasting and microstructure noise. Available at URL: www.unc.edu~eghysels.

Ghysels, Eric, Sinko, Arthur, Valkanov, Rossen, 2007. MIDAS regressions: further results and new directions. Econometric Reviews 26, 53–90.

Granger, C.W.J., Newbold, P., 1976. Forecasting transformed series. Journal of the Royal Statistical Society. Series B (Methodological) 189–203.

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

Meddahi, Nour, 2001. An eigenfunction approach for volatility modeling. Working Paper. University of Montreal.

Nelson, Daniel B., 1990. ARCH models as diffusion approximations. Journal of Econometrics 45, 7–38.

Oomen, R.A.C., 2005. Properties of bias corrected realized variance in calendar time and business time. Journal of Financial Econometrics 3 (4), 555–577.

Renault, E., 2009. Moment-based estimation of stochastic volatility models. In: Andersen, T.G., Davis, R.A., Kreiss, J., Mikosch, T. (Eds.), Handbook of Financial Time Series. Springer Verlag, Berlin.

Sinko, Arthur, 2007. On predictability of market microstructure noise volatility. UNC Working Paper.

Taylor, S.J., 1982. Financial returns modelled by the product of two stochastic processes. A study of daily sugar prices, 1961–79. In: Shephard, N. (Ed.), Stochastic Volatility: Selected Readings. Oxford University Press, Oxford.

Wasserfallen, Walter, Zimmermann, Heinz, 1985. The behavior of intraday exchange rates. Journal of Banking and Finance 9, 55–72.

Zhang, Lan, Mykland, Per A., Aït-Sahalia, Yacine, 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, Bin, 1996. High-frequency data and volatility in foreign-exchange rates. Journal of Business and Economic Statistics 14, 45–52.