In-sample Asymptotics and Across-sample Efficiency Gains for High Frequency Data Statistics
Online Appendix

Eric Ghysels* **Per Mykland†** **Eric Renault‡**

June 8, 2021

Abstract

We revisit in-sample asymptotic analysis extensively used in the realized volatility literature. We show that there are gains to be made in estimating current realized volatility from considering realizations in prior periods. The weighting schemes also relate to Kalman-Bucy filters, although our approach is non-Gaussian and model-free. We derive theoretical results for a broad class of processes pertaining to volatility, higher moments and leverage. The paper also contains a Monte Carlo simulation study showing the benefits of across-sample combinations.

*University of North Carolina, Department of Economics, Gardner Hall CB 3305, Chapel Hill, NC 27599-3305, e-mail: eghysels@unc.edu.

†University of Chicago, Department of Statistics, 5734 University Avenue, Chicago, Illinois 60637, e-mail: mykland@galton.uchicago.edu.

‡Department of Economics, University of Warwick, Coventry, CV4 7AL, United Kingdom, e-mail: Eric.Renault@warwick.ac.uk.

In this Online Appendix we provide technical details pertaining to the results in Ghysels, Mykland, and Renault (2021). Section OA.1 covers the proof of Proposition 3.1. Additional volatility examples are covered in Section OA.2, whereas examples beyond volatility appear in Section OA.3. Section OA.4 provides insights on the connection of our analysis with the Kalman filter. Finally, Section OA.5 discusses a forecasting example.

OA.1 Proof of Proposition 3.1

We study the details for across-sample efficiency gains pertaining to the first Example 1 for the standard realized volatility (RV), obtained by summing squared intra-daily returns, yielding the so called realized variance, namely:

$$\hat{\Theta}_{n,t} = \sum_{t-1 < t_{i+1} \leq t} (X_{t_{i+1}} - X_{t_i})^2. \tag{OA.1.1}$$

We study here the case of equidistant sampling, $t_i - t_{i-1} = \Delta\ t_i = 1/n$. When the sampling frequency increases, i.e. $n \to \infty$, then the realized variance converges uniformly in probability to the increment of the quadratic variation i.e.

$$\lim_{n\to\infty} \hat{\Theta}_{n,t} \to^p \int_{t-1}^{t} \theta(s)ds. \tag{OA.1.2}$$

To streamline the notation we will drop the superscript n. Barndorff-Nielsen and Shephard (2002), Jacod and Protter (1998) and Zhang (2001) show that the error of realized variance is asymptotically

$$\frac{n^{1/2}(\hat{\Theta}_{n,t} - \int_{t-1}^{t} \theta(s)ds)}{\sqrt{V_t}} \xrightarrow{d} N(0,1) \tag{OA.1.3}$$

where $V_t = 2\int_{t-1}^{t} \sigma(s)^4 ds$ is called (twice) the integrated quarticity. It should be noted that in the case of no leverage effect, the result in (OA.1.3) follows directly from the simplified example in Section 2 of Ghysels, Mykland, and Renault (2021).

We want to estimate $\Theta_t = \int_{t-1}^{t} \sigma^2(s)ds$ and take advantage of observations on the previous day, summarized by $\hat{\Theta}_{n,t-1}$, the estimator of Θ_{t-1}. The key assumption is that the two estimators $\hat{\Theta}_{n,\tau}$, $\tau = t-1$ and t have an asymptotic accuracy of the same order of magnitude and are asymptotically independent, for a given volatility path, namely:

$$\begin{aligned} \frac{n^{1/2}(\hat{\Theta}_{n,t-1} - \Theta_{t-1})}{\sqrt{V_{t-1}}} &\xrightarrow{d} N(0,1) \\ \frac{n^{1/2}(\hat{\Theta}_{n,t} - \Theta_{t})}{\sqrt{V_{t}}} &\xrightarrow{d} N(0,1) \end{aligned} \tag{OA.1.4}$$

and the joint asymptotic distribution is the product of the marginals. We consider possible improvements of our estimator of $\int_{t-1}^{t} \theta(s)ds$, assuming for the moment that we know the correlation coefficient:

$$\varphi_n = \frac{Cov(\hat{\Theta}_{n,t}, \hat{\Theta}_{n,t-1})}{Var(\hat{\Theta}_{n,t-1})} \tag{OA.1.5}$$

and the unconditional expectation $E(\hat{\Theta}_{n,t}) = E(\hat{\Theta}_{n,t-1}) = E(\int_{t-1}^{t} \sigma^2(s)ds)$. Note that equation (OA.1.5) does *not* imply that our analysis is confined to AR(1) models. Instead, equation (OA.1.5) only reflects the fact that we condition predictions on a single lag $\hat{\Theta}_{t-1}$. Equation (OA.1.5) is essentially a general version of equation (2.6). There may be potential gains from considering more lags, as the underlying models would result in higher order dynamics. Yet, for our analysis we currently focus exclusively on prediction equations with a single lag. Higher order equations are a straightforward extension discussed later. In analogy with equation (2.12) we also need:

$$\varphi_n^0 = \frac{Cov(\Theta_{n,t}, \Theta_{n,t-1})}{Var(\Theta_{n,t-1})}. \tag{OA.1.6}$$

The theory presented in the sequel will mirror the development in Section 2 of Ghysels, Mykland, and Renault (2021), but be valid even when volatility is not piecewise constant.

Consider the best linear forecast of using (only) $\hat{\Theta}_{n,t-1}$:

$$\breve{\Theta}_{n,t|t-1} = \varphi_n \hat{\Theta}_{n,t-1} + (1 - \varphi_n) E\left(\Theta_t\right). \tag{OA.1.7}$$

Note that this *realized forecast* is infeasible in practice and, to make it feasible, estimators of φ and $E(\int_{t-1}^{t} \sigma^2(s)ds)$ are required. These estimators will be based on past time series of realized volatilities: $\hat{\Theta}_{n,\tau}$, $\tau = t-1, ..., t-T+1$. The estimation error on these coefficients will be made negligible when (T/n) goes to infinity.

Our goal is to combine the two measurements $\hat{\Theta}_{n,t}$ and $\breve{\Theta}_{n,t|t-1}$ of $\int_{t-1}^{t} \theta(s)ds$ to define a new estimator:

$$\hat{\Theta}_{n,t}(\omega_t) = (1-\omega_t)\hat{\Theta}_{n,t} + \omega_t \breve{\Theta}_{n,t|t-1}. \tag{OA.1.8}$$

Intuitively, the more persistent the volatility process, the more $\hat{\Theta}_{n,t|t-1}$ is informative about $\hat{\Theta}_{n,t}$ and the larger the optimal weight ω_t should be. Note that the weight depends on t, as indeed its computation will be volatility path dependent. To characterize such an optimal choice, one may apply a conditional control variable principle, given the volatility path. The criterion to minimize will be the conditional mean squared error:

$$E_\sigma \left[\hat{\Theta}_{n,t}(\omega_t) - \Theta_t\right]^2 = E_\sigma \left\{\hat{\Theta}_{n,t} - \Theta_t - \omega_t \left(\hat{\Theta}_{n,t} - \breve{\Theta}_{n,t|t-1}\right)\right\}^2. \tag{OA.1.9}$$

Then, the problem to solve is obviously nearly identical to the one considered in Section 2 of Ghysels, Mykland, and Renault (2021), so that:

$$\hat{\Theta}_{n,t}(\omega^*_{n,t}) = \hat{\Theta}_{n,t} - \omega^*_{n,t}\left(\hat{\Theta}_{n,t} - \breve{\Theta}_{n,t|t-1}\right) \tag{OA.1.10}$$

will be an optimal improvement of $\hat{\Theta}_{n,t}$ if $\omega^*_{n,t}$ is defined according to the following control variable formula:

$$\omega^*_{n,t} = \frac{Cov_\Theta\left[\hat{\Theta}_{n,t}, \hat{\Theta}_{n,t} - \breve{\Theta}_{n,t|t-1}\right]}{E_\Theta\left[\hat{\Theta}_{n,t} - \breve{\Theta}_{n,t|t-1}\right]^2} = \frac{Var_\Theta[\hat{\Theta}_{n,t}]}{E_\Theta\left[\hat{\Theta}_{n,t} - \breve{\Theta}_{n,t|t-1}\right]^2}. \tag{OA.1.11}$$

Note that $\omega^*_{n,t}$ has been shrunk with respect to the conditional regression coefficient $\hat{\Theta}_{n,t}$ on ($\hat{\Theta}_{n,t}$ - $\breve{\Theta}_{n,t|t-1}$). This is due to the need to take into account the non-zero mean of $(\hat{\Theta}_{n,t} - \breve{\Theta}_{n,t|t-1})$ given the volatility path.

A closed form formula for the optimal weights is obtained by computing moments, given the volatility path, according to the asymptotic distribution appearing in (OA.1.4). Then, given the volatility path, we have:

$$\begin{aligned}
E_\sigma\left(\hat{\Theta}_{n,t} - \breve{\Theta}_{n,t|t-1}\right) &= \Theta_t - \varphi_n\Theta_{t-1} - (1-\varphi_n)E\left[\Theta_t\right] \\
Var_\sigma\left(\hat{\Theta}_{n,t} - \hat{\Theta}_{n,t|t-1}\right) &= \frac{V_t}{n} + \varphi_n^2\frac{V_{t-1}}{n} + o\left(\frac{1}{n}\right) \\
Cov_\sigma\left[\hat{\Theta}_{n,t}, \hat{\Theta}_{n,t} - \hat{\Theta}_{n,t|t-1}\right] &= Var_\sigma\left[\hat{\Theta}_{n,t}\right] = \frac{V_t}{n} + o\left(\frac{1}{n}\right).
\end{aligned}$$

Therefore, $\omega_{n,t}^*$ defined by (OA.1.10) can be described by:

$$\begin{aligned} \frac{1}{\omega_{n,t}^*} &= 1+\varphi_n^2\frac{V_{t-1}}{V_t}+n\frac{B_{F,n}^2(t)}{V_t}+o\left(\frac{1}{n}\right) \\ B_{F,n}(t) &= \Theta_t-\varphi_n\Theta_{t-1}-(1-\varphi_n)E\left[\Theta_t\right]. \end{aligned} \tag{OA.1.12}$$

The above yields the result in equation (3.7) of Proposition 3.1. Note that, in order to understand this optimal weight, it is useful to rewrite it as follows (as a general version of (2.11) in Ghysels, Mykland, and Renault (2021)):

$$\begin{aligned} \frac{1}{\omega_{n,t}^*} &= 1+\varphi_n^2\frac{V_{t-1}}{V_t}+n\frac{B_{F,n}^2(t)-B_I^2(t)}{V_t}+n\frac{B_I^2(t)}{V_t}+o\left(\frac{1}{n}\right) \\ B_I(t) &= \Theta_t-\varphi^0\Theta_{t-1}-(1-\varphi^0)E\left[\Theta_t\right] \\ \varphi^0 &= \frac{Cov\left[\Theta_t,\Theta_{t-1}\right]}{Var\left(\Theta_t\right)}=\varphi_n+O\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

Moving to Example 2, in order to separate the jump and continuous sample path components in quadratic variation, (and thus estimate $\int_{t-1}^t\sigma^2(s)ds$), Barndorff-Nielsen and Shephard (2004a) and Barndorff-Nielsen and Shephard (2004b)introduce the concept of bi-power variation (BPV) defined as:

$$\hat{\Theta}_{n,t}(k)=\frac{\pi}{2}\sum_{j=k+1}^{n}\left|X_{n,t_j}\right|\left|X_n,t_{j-k}\right|. \tag{OA.1.13}$$

Henceforth we will, without loss of generality, specialize our discussion to the case $k=1$, and therefore drop it to simplify notation. Barndorff-Nielsen and Shephard (2004b) establish the sampling behavior of $\hat{\Theta}_{n,t}$ as $n\to\infty$, and show that under suitable regularity conditions: $\text{plim}_{n\to\infty}\hat{\Theta}_{n,t}(k)=\int_{t-1}^t\sigma(s)^2ds$. Therefore, in the presence of jumps, $\hat{\Theta}_{n,t}$ converges to the continuous path component of $\int_{t-1}^t\theta(s)ds$ and is not affected by jumps. The sampling error of the bi-power variation is $n^{\alpha}\left(\hat{\Theta}_{n,t}-\int_{t-1}^t\sigma\left(s\right)^2ds\right)/\sqrt{\nu_{bb}V_t}\sim N\left(0,1\right)$, where $\nu_{bb}\approx 2.6$ and under the null where there are no jumps. Based on these results, Barndorff-Nielsen and Shephard (2004a) and Barndorff-Nielsen and Shephard (2004b) introduce a framework to test for jumps based on the fact that QV consistently estimates the quadratic variation, while $\hat{\Theta}$ consistently estimates the integrated variance, even in the presence of jumps. Therefore, the difference between the BPV and the scaled RV is sum of squared jumps (in the limit).

Once we have identified the jump component, we can subtract it from the realized variance and we will have the continuous part of the process.

Using the arguments presented earlier we can improve estimates of both RV and BPV. This should allow us to improve estimates of integrated volatility as well as improve the performance of tests for jumps. To do so we introduce:

$$\hat{\Theta}_{n,t}(\omega_t^*) = \hat{\Theta}_{n,t} - \omega_t^*(\hat{\Theta}_{n,t} - \hat{\Theta}_{n,t|t-1}). \tag{OA.1.14}$$

It will be an optimal improvement of $\hat{\Theta}_{n,t}$ when ω_t^* is again defined according to the control variable formula (OA.1.11) where QV is replaced by BPV. Note that we do not assume the same temporal dependence for QV and BPV, as the projection of QV on its past (one lag) and that of BPV on its own past (one lag) in general do not coincide.

OA.2 Additional Volatility Examples

Example 1 (Two-Scales Realized Volatility) The observations are as in (3.2) of the paper. The classical Two-Scales Realized Volatility (TSRV; Zhang, Mykland, and Aït-Sahalia (2005), Aït-Sahalia, Mykland, and Zhang (2011)) has a convergence rate of $\alpha = 1/6$. The cited papers show consistency and stable convergence. □

Example 2 (Multi-Scale and Kernel Realized Volatility) The observations are again as in (3.2). The convergence rate is $\alpha = 1/4$. Conditions [i] and [iia] (and in particular [ii]) have been shown for Multi-Scale Realized Volatility (MSRV, Zhang (2006)) and Realized Kernel estimators (RK, Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008)). The two methods are close in view of Bibinger and Mykland (2016). The rate is efficient since it also occurs in the parametric case (see Gloter and Jacod (2000)), and also see the non-parametric bound in Reiss (2011). □

In the interest of brevity we discuss the above Example 1 in more detail. The efficient price X_t is now latent and observed with microstructure noise, as in equation (3.2) of the paper. The analysis in the earlier sections for the no-jump case goes through with these modifications. In the case of the TSRV,

$$V_t \;=\; c\frac{4}{3}\int_{t-1}^{t} \sigma_u^4 du + 8c^{-2}\nu^4,$$

where ν^2 is the variance of the noise, and c is a smoothing parameter. For the more complicated case of the MSRV and the kernel estimators, we refer to the publications cited.[1] In the case where there are both jumps and microstructure, there are two different targets that can be considered, either the full quadratic variation Θ_t, or only its continuous part. For estimation of the full quadratic variation, the estimators from the continuous case remain consistent, and retain the same rate of convergence as before. The asymptotic variance V_t needs to be modified. The results in this paper for the no-jump case therefore remain valid.[2]

For other approaches to the estimation of volatility under microstructure noise, see, *inter alia*, Aït-Sahalia, Mykland, and Zhang (2005), Bandi and Russell (2006) and Hansen and Lunde (2006), Jacod, Li, Mykland, Podolskij, and Vetter (2009), Podolskij and Vetter (2009b), Xiu (2010), and Bibinger and Reiss (2014). The case of bi-power variation can be also extended to measures involving more general functions, as in Barndorff-Nielsen, Graversen, Jacod, and Shephard (2006). Provided such measure feature persistence we can apply the above analysis in a more general context. One particular case of interest is power variation, typically more persistent than quadratic variation or related measures, as discussed in detail in Forsberg and Ghysels (2006).

OA.3 Beyond Realized Volatility

The set of examples discussed in this section go beyond measures of quadratic variation.

[1]For estimation of the continuous part of Θ_t only, apart from bi- and multipower, the most fully developed theory involves removal by truncation, see Mancini (2001), Fan and Wang (2007), Aït-Sahalia and Jacod (2007), Aït-Sahalia and Jacod (2008), Aït-Sahalia and Jacod (2009), Aït-Sahalia and Jacod (2012), Jacod and Todorov (2010), Lee and Mykland (2008), Lee and Mykland (2012), Jing, Kong, Liu, and Mykland (2012), as well as the work cited in the previous paragraph. The work by Huang and Tauchen (2006) provides a complete theory, but under the assumption that the microstructure noise is Gaussian. Irregular observations can be handled using the concept of quadratic variation of time (see e.g. Mykland and Zhang (2006), Mykland and Zhang (2016), Mykland, Zhang, and Chen (2019)).

[2]See, Jacod and Protter (2012), Aït-Sahalia and Jacod (2014), Mykland and Zhang (2016), Mykland, Zhang, and Chen (2019).

Example 3 (Estimation of Covariance from Asynchronous Observations) There are a number of different ways of handling covariances, including the Hayashi and Yoshida (2005) estimator, see also Podolskij and Vetter (2009a), Christensen, Podolskij, and Vetter (2013), and Bibinger and Vetter (2015) for micro-structure, jumps, and asymptotic distributions. Alternatives include the Previous-Tick estimator (Zhang (2011), Bibinger and Mykland (2016)), and Quasi-Likelihood (Shephard and Xiu (2012)). The estimator in Mykland and Zhang (2012, p. 172-175) is a hybrid of Hayashi-Yoshida and Quasi-Likelihood. Consistency and stable convergence hold here again with varying rates α. □

Example 4 (Block Estimation of Higher Powers of Volatility) The parameter of interest is $\theta(s) = g(\sigma(s)^2)$, with g not being the identity function. To make estimators approximately or fully efficient, one can use block estimation. In the absence of microstructure noise, the convergence rate is $\alpha = 1/2$ (Mykland and Zhang (2009, p. 1421-1426)), Mykland and Zhang (2011, p. 224-229), Jacod and Rosenbaum (2015) and Jacod and Rosenbaum (2013)). If microstructure noise is present, the convergence rate is $\alpha = 1/4$ (Jacod and Protter (2012, p. 512-554)).[3] □

Next, we are concerned with systems on the form $dV_t = \beta_t dX_t + dZ_t$, where V_t and X_t can be observed at high frequency, either with or without microstructure noise. Moreover, X_t can be multidimensional. The coefficient process β_t can either be the "beta" from portfolio optimization, with Z_t in the role of idiosyncratic noise, or β_t can be the hedging "delta" for an option, with Z_t as tracking error. Nonparametric estimates can be used directly, or for forecasting, or for model checking.

Example 5 (High Frequency Regression, and ANOVA) The regression problem seeks to estimate or make tests about $\int_{t-1}^{t} \beta(s)ds$ (Mykland and Zhang (2009, p. 1424-1426), Kalnina (2012), Zhang (2012, p. 268-273), Reiss, Todorov, and Tauchen (2015)). The ANOVA problem seeks to f $[Z, Z]_t$ (Zhang (2001) and Mykland and Zhang (2006)). Convergence rates are as for realized or other powers of volatility, with $\alpha = 1/2$ when there is no microstructure noise, and $\alpha = 1/4$ otherwise. □

[3] See also Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006), Mykland and Zhang (2012, p. 138) and Renault, Sarisoy, and Werker (2017) for related developments.

OA.4 Connection with Kalman Filter

In-sample asymptotics applies to situations where the period t variable of interest, say Y_t, is observed by a sequence of noisy measurements X_t which are, when the number n of intraday observations goes to infinity, asymptotically unbiased and normally distributed with a possibly time dependent and random variance. In other words, adopting a commonly used abuse of notation, we have for large n a measurement equation:

$$X_t = Y_t + c_t Z_t \qquad Z_t \sim N(0,1). \tag{OA.4.1}$$

Traditional Kalman filtering applies to a Gaussian measurement equation like (OA.4.1) augmented with a Gaussian transition equation:

$$Y_t = aY_{t-1} + g + \sqrt{H_t} u_t \qquad u_t \sim N(0,1). \tag{OA.4.2}$$

Note that the Kalman filter considered here is slightly more general than the traditional one, since we allow for conditional heteroskedasticity in equations (OA.4.1) and (OA.4.2). Conditioning on the values of the variables c_t and H_t allows us to consider heteroskedasticity as predetermined and to use standard Bayes formula to obtain recursion formulas for the filtering distribution, that is the conditional normal distribution of Y_t given $(X_t, X_{t-1}, \ldots, X_1)$. If $N(Y_t^*, K_t)$ stands for this conditional distribution (the superscript n is omitted for convenience), the Kalman recursion formulas (see e.g. Williams (1991), p. 168) can be written:

$$\begin{aligned} \frac{1}{K_t} &= \frac{1}{a^2 K_{t-1} + H_t} + \frac{1}{c_t^2} \\ \frac{Y_t^*}{K_t} &= \frac{aY_{t-1}^* + g}{a^2 K_{t-1} + H_t} + \frac{X_t}{c_t^2}. \end{aligned} \tag{OA.4.3}$$

Hence, the filtered values are defined as:

$$Y_t^* = (1 - \omega_t) X_t + \omega_t (aY_{t-1}^* + g) \tag{OA.4.4}$$

with:

$$\omega_t = \frac{c_t^2}{c_t^2 + a^2 K_{t-1} + H_t}. \tag{OA.4.5}$$

As will be discussed in detail later, our way to take advantage of past information to improve upon the measurement X_t is quite similar in spirit to the Kalman-Bucy filter with however a couple of important differences. First, we do not deal with a Gaussian transition equation. The processes we consider, like realized variance or bi-power variation (defined later) are often constrained to be positive (and therefore cannot be Gaussian). However, even more importantly, we want to remain model free. Therefore, instead of the model (OA.4.2), we will simply refer to a linear projection of the variable Y_t on some of its lagged values. $Y_{t|t-1}$ will denote such an optimal linear predictor for a given number of lags. If for instance we decide to use only one lag, α and g will be defined as (maintaining a model-free setting):

$$\begin{aligned} Y_{t|t-1} &= aY_{t-1} + g \\ Y_t &= Y_{t|t-1} + u_t \\ E(u_t) &= 0, \qquad Cov(u_t, Y_{t-1}) = 0. \end{aligned}$$

Note that the structure of equation (OA.4.1) is maintained but in a model-free context, up to some regularity conditions ensuring its asymptotic validity.

Second, without a Gaussian transition equation we have to give up the Bayes formula which enables us to compute conditional expectations and variances like (OA.4.3). Moreover, maintaining the model-free setting, we do not want to define a filtered value function of *all* past information $(X_t, X_{t-1}, ..., X_1)$, but instead only for a fixed number of lags every single day t. For instance, in the case of one lag, our filtered value will be defined by:

$$\tilde{Y}_t = (1 - \omega_t^*)X_t + \omega_t^*(aX_{t-1} + g). \tag{OA.4.6}$$

Note that the actual filtered value will be slightly different due to estimation error, an issue that also appears in traditional Kalman filtering and that will be discussed later. It is important to note that our filtered value $\tilde{Y}_t$ is different from the Kalman filter Y_t^* in two respects. First, it combines the current measurement X_t with the measured optimal forecast $aX_{t-1} + g$ based on yesterday's information (measured counterpart of $Y_{t|t-1} = aY_{t-1} + g$) and not on the filtered counterpart $aY_{t-1}^* + g$. Second, we have shown that an optimal weight for minimizing the conditional mean squared error is given by:

$$\omega_t^* = \frac{c_t^2}{c_t^2 + a^2 c_{t-1}^2 + H_t}. \tag{OA.4.7}$$

Note that our optimal weight (OA.4.7), albeit similar to the Kalman one (OA.4.5), is smaller since $c_t^2 > K_t$ (using equation (OA.4.3)). Intuitively, our weighting schemes give less weight to past information since we summarize past information by the measured counterpart $aX_{t-1} + g$ (with conditional variance $a^2c_{t-1}^2$) of the past forecast $Y_{t|t-1} = aY_{t-1} + g$, whereas the Kalman filter weights use the more accurate filtered counterpart $aY_{t-1}^* + g$ (i.e. with smaller conditional variance $a^2K_{t-1} < a^2c_{t-1}^2$).

OA.5 A Forecasting Example

We mainly focus on the *measurement* of high frequency data related processes such as quadratic variation, bi-power variation and quarticity. Yet, forecasting future realizations of such processes is often the ultimate goal. The purpose of this subsection is to discuss the impact on forecasting performance of our improved measurements.

We start from the observation that standard volatility measurements feature a measurement error that can be considered, at least asymptotically, as a martingale difference sequence. Therefore, suppose we want to forecast Y_{t+1}, using past observations $(X_s), s \leq t$ which are noisy measurements of past $Y's$. The maintained martingale difference assumption implies that:

$$Cov[Y_t - X_t, X_s] = 0, \forall s < t.$$

Suppose now that we also consider past "improved" observations: $Z(s) = (1-\omega)X_s + \omega Y(s)^*$, where $Y(s)^*$ is an unbiased linear predictor of X_s.[4] Since Y_{t+1}^* is an unbiased linear predictor of X_{t+1} :

$$X_{t+1} = Y_{t+1}^* + v_{t+1}^*, E(v_{t+1}^*) = 0, Cov[v_{t+1}^*, Y_{t+1}^*] = 0.$$

Suppose we have a preferred forecasting rule for (X_t) (based on say an ARFIMA model for QV such as in Andersen, Bollerslev, Diebold, and Labys (2003)) and let us denote it as Y_{t+1}^X. This alternative unbiased linear predictor of X_{t+1} would be such that:

$$X_{t+1} = Y_{t+1}^X + v_{t+1}^X, E(v_{t+1}^X) = 0, Cov[v_{t+1}^X, Y_{t+1}^X] = 0.$$

[4] We consider here weights that are not time varying - unlike in the previous subsection. Hence, we are assessing here the impact on forecasting performance of fixed weights ω. Optimally chosen time varying weights should ensure at least a comparable forecasting performance. We will refer to the latter as conditional schemes, in contrast to unconditional ones that are also discussed in subsection 4 of the paper.

It is natural to assume in addition that:

$$Cov[v_{t+1}^X, Y_{t+1}^*] = 0$$

hence, the predictor Y_{t+1}^* does not allow us to improve the preferred predictor Y_{t+1}^X. Consider now a modified forecast Y_{t+1}^Z, based on past and current improved observations $Z(s), s \leq t$. By the definition of Z we have:

$$Y_{t+1}^Z = (1-\omega)Y_{t+1}^X + \omega Y_{t+1}^*.$$

It is easy to show that the forecasting errors obtained from respectively Y_{t+1}^X and Y_{t+1}^Z satisfy:

$$Var(Y_{t+1}^X - Y_{t+1}) - Var(Y_{t+1}^Z - Y_{t+1}) = \omega^2(Var(v_{t+1}^X) - Var(v_{t+1}^*)). \tag{OA.5.8}$$

This result has the following implication: using the proxy (Z) instead of the proxy (X) we will not deteriorate the forecasting performance, except if we build on purpose the proxy (Z) from a predictor (Y^*) less accurate than the preferred predictor (Y^X). More generally, using a proxy Z computed with time varying weights optimally chosen in a conditional setting should indeed improve the forecasting performance.

The above result pertains to a simple linear forecasting setting. We expect that improvements of measurement are going to be even more important nonlinear settings. The most prominent example where the objective of interest is a nonlinear function of future volatility is option pricing. The future path of volatility until some time to maturity of the derivative contract determines the current option price through a conditional expectation of a nonlinear payoff function. A simplified example is the model of Hull and White (2005) where the price of a European call option of time to maturity h and moneyness k equals:

$$C_t(h,k) = E_t[BS(\frac{1}{h}\int_t^{t+h} \sigma^2(u)du, k, h)]$$

and $BS(\sigma^2, k, h)$ is the Black-Scholes option price formula. Note that the above equation assumes no leverage and no price of volatility risk.[5] This is a common example of derivative pricing that will be studied later via simulation. It will be shown that the improved volatility measurement has a significant impact on option pricing.

[5] Note that volatility is not priced in Hull and White (2005), but nevertheless it is still a conceptually interesting example.

The conclusions regarding the ranking of weights appearing in Section 5 of the paper change when we turn our attention to forecasting gains associated with the example discussed in this section. We consider two experiments. The first is a linear forecasting exercise:

$$IV_{t+1} = a + bX_t + \varepsilon_{t+1}^X$$

where the regressor X_t is either one of the following: IV_t, $\hat{\Theta}_t$ and corrected $\hat{\Theta}_{n,t}$ using the optimal weighting ω_{vt}^* and ω_{ut}^* schemes with one-day lag of information. Obviously the infeasible benchmark is the regression involving IV as regressor. Hence, we compare how close the feasible raw $\hat{\Theta}_{n,t}$ and corrected measure perform in comparison. We do this for 5 and 10 minute sampling schemes.

To appraise the more realistic and interesting nonlinear forecasting setting we consider the following prediction problem:

$$\log(BS_{t+k}^{imp}(ATM, TTM)) = a + b\log(X_t) + \varepsilon_{kt}^{X,TTM}$$

where k = 1 day, 5 days, 20 days. The $BS_{t+k}^{imp}(ATM, TTM)$ is the Black-Scholes implied volatility generated for a sample of data obeying the stochastic volatility dynamics of the Heston model mentioned in the first subsection. We selected the Heston because we know how to price options, and hence compute Black-Scholes implied volatilities. We picked three times-to-maturity (TTM), namely 22, 44 and 66 days, corresponding to one-, two- and three-month options. Finally, we focused exclusively on at-the-money options (ATM) since those are typically accurately priced and liquid.

The simulation evidence is reported in Table OA.1 and focuses exclusively on the second experiment described above - as the first one did not yield many differences between the various estimation schemes. There is now a clear difference between the ranking in Table 2 in the main paper and that in Table OA.1. In the case of filtering, we find that ω_{vt}^* is slightly better than ω_{ut}^*. Hence, the conditional weighting scheme with realized U_t, i.e. ω_{ut}^* outperforms ω_{vt}^*, and therefore as far as forecasting a nonlinear function goes, the ranking is reversed. Although the differences are not as significant, there is a clear pattern of dominance which we did not observe in Table 2 pertaining to volatility appearing in the paper.

Table OA.1: Forecasting Log Black-Scholes Implied Volatilities - Heston Model

We consider:

$$\log(BS^{imp}_{t+k}(ATM, TTM)) = a + b\log(X_t) + \varepsilon^{X,TTM}_{kt}$$

where k = 1 day, 5 days, 20 days. The $BS^{imp}_{t+k}(ATM, TTM)$ is the Black-Scholes implied volatility generated by a Heston model for at-the-money options (ATM) with times-to-maturity (TTM) 22, 44 and 66 days.

X_t		IV	QV Corrected(ω^*_{vt})	QV Corrected(ω^*_{ut})	QV
		R^2 1-day ahead forecast, ATM			
TTM: 22days	5 Min	0.98	0.93	0.94	0.93
	10 Min	0.98	0.88	0.91	0.90
TTM: 44days	5 Min	0.98	0.91	0.94	0.93
	10 Min	0.98	0.86	0.91	0.90
TTM: 66days	5 Min	0.98	0.93	0.94	0.93
	10 Min	0.98	0.92	0.94	0.93
		R^2 5-day ahead forecast, ATM			
TTM: 22days	5 Min	0.91	0.86	0.88	0.87
	10 Min	0.91	0.82	0.85	0.84
TTM: 44days	5 Min	0.91	0.86	0.88	0.87
	10 Min	0.91	0.82	0.85	0.84
TTM: 66days	5 Min	0.91	0.85	0.87	0.87
	10 Min	0.91	0.82	0.85	0.84
		R^2 20-day ahead forecast, ATM			
TTM: 22days	5 Min	0.72	0.68	0.70	0.69
	10 Min	0.72	0.65	0.67	0.66
TTM: 44days	5 Min	0.72	0.68	0.69	0.69
	10 Min	0.72	0.65	0.67	0.67
TTM: 66days	5 Min	0.72	0.68	0.69	0.69
	10 Min	0.72	0.64	0.67	0.67

References

Aït-Sahalia, Y., and J. Jacod (2007): "Volatility Estimators for Discretely Sampled Lévy Processes," *Annals of Statistics*, 35, 335–392.

——— (2008): "Fisher's Information for Discretely Sampled Lévy Processes," *Econometrica*, 76, 727–761.

——— (2009): "Testing for jumps in a discretely observed process," *Annals of Statistics*, 37, 184–222.

——— (2012): "Analyzing the Spectrum of Asset Returns: Jump and Volatility Components in High Frequency Data," *Journal of Economic Literature*, 50, 1007–1150.

——— (2014): *High-Frequency Financial Econometrics*. Princeton University Press, Princeton, N.J.

Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (2005): "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise," *Review of Financial Studies*, 18, 351–416.

——— (2011): "Ultra high frequency volatility estimation with dependent microstructure noise," *Journal of Econometrics*, 160, 160–175.

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

Bandi, F. M., and J. R. Russell (2006): "Separating microstructure noise from volatility," *Journal of Financial Economics*, 79, 655–692.

Barndorff-Nielsen, O., S. Graversen, J. Jacod, M. Podolskij, and N. Shephard (2006): "A central limit theorem for realised power and bipower variations of continuous semimartingales," in *From Stochastic Calculus to Mathematical Finance, The Shiryaev Festschrift*, ed. by Y. Kabanov, R. Liptser, and J. Stoyanov, pp. 33–69. Springer Verlag, Berlin.

Barndorff-Nielsen, O., S. Graversen, J. Jacod, and N. Shephard (2006): "Limit theorems for bipower variation in financial econometrics," *Econometric Theory*, 22, 677–719.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008): "Designing realised kernels to measure ex-post variation of equity prices in the presence of noise," *Econometrica*, 76(6), 1481–1536.

Barndorff-Nielsen, O. E., and N. Shephard (2002): "Estimating Quadratic Variation Using Realised Variance," *Journal of Applied Econometrics*, 17, 457–477.

——— (2004a): "Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation," *Journal of Financial Econometrics*, 4, 1–30.

——— (2004b): "Power and Bipower Variation with Stochastic Volatility and Jumps," *Journal of Financial Econometrics*, 2, 1–48.

Bibinger, M., and P. A. Mykland (2016): "Inference for Multi-dimensional High-frequency Data with an Application to Conditional Independence Testing," *Scandinavian Journal of Statistics*, 43(4), 1078–1102.

Bibinger, M., and M. Reiss (2014): "Spectral Estimation of Covolatility from Noisy Observations Using Local Weights," *Scandinavian Journal of Statistics*, 41(1), 23–50.

Bibinger, M., and M. Vetter (2015): "Estimating the quadratic covariation of an asynchronously observed semimartingale with jumps," *Annals of the Institute of Statistical Mathematics*, 67, 707–743.

Christensen, K., M. Podolskij, and M. Vetter (2013): "On covariation estimation for multivariate continuous Itô semimartingales with noise in non-synchronous observation schemes," *Journal of Multivariate Analysis*, 120, 59–84.

Fan, J., and Y. Wang (2007): "Multi-scale Jump and Volatility Analysis for High-Frequency Financial Data," *Journal of the American Statistical Association*, 102, 1349–1362.

Forsberg, L., and E. Ghysels (2006): "Why do absolute returns predict volatility so well?," *Journal of Financial Econometrics*, 4, 31–67.

Ghysels, E., P. Mykland, and E. Renault (2021): "In-sample Asymptotics and Across-sample Efficiency Gains for High Frequency Data Statistics," *Econometric Theory*, Conditionally Accepted.

Gloter, A., and J. Jacod (2000): "Diffusions with Measurement Errors: I - Local Asymptotic Normality and II - Optimal Estimators," Discussion paper, Université de Paris-6.

Hansen, P. R., and A. Lunde (2006): "Realized Variance and Market Microstructure Noise," *Journal of Business and Economic Statistics*, 24, 127–161.

Hayashi, T., and N. Yoshida (2005): “On Covariance Estimation of Non-synchronously Observed Diffusion Processes,” *Bernoulli*, 11, 359–379.

Huang, X., and G. Tauchen (2006): “The relative contribution of jumps to total price variance,” *Journal of Financial Econometrics*, 4, 456–499.

Hull, J., and A. White (2005): “The Pricing of Options on Assets with Stochastic Volatilities,” *Stochastic Volatility: Selected Readings*.

Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter (2009): “Microstructure Noise in the Continuous Case: The Pre-Averaging Approach,” *Stochastic Processes and Their Applications*, 119, 2249–2276.

Jacod, J., and P. Protter (1998): “Asymptotic Error Distributions for the Euler Method for Stochastic Differential Equations,” *Annals of Probability*, 26, 267–307.

Jacod, J., and P. Protter (2012): *Discretization of Processes*. Springer-Verlag, New York, first edn.

Jacod, J., and M. Rosenbaum (2013): “Quarticity and other Functionals of Volatility: Efficient Estimation,” *Annals of Statistics*, 41, 1462–1484.

——— (2015): “Estimation of Volatility Functionals: The Case of a $\sqrt{n}$ Window,” in *Large Deviations and Asymptotic Methods in Finance*, pp. 559–590. Springer.

Jacod, J., and V. Todorov (2010): “Do Price and Volatility Jump Together?,” *Annals of Applied Probability*, 20 (4), 1425–1469.

Jing, B.-Y., X.-B. Kong, Z. Liu, and P. A. Mykland (2012): “On the jump activity index for semimartingales,” *Journal of Econometrics*, 166, 213–223.

Kalnina, I. (2012): “Nonparametric Tests of Time Variation in Betas,” discussion paper.

Lee, S. S., and P. A. Mykland (2008): “Jumps in Financial Markets: A New Nonparametric Test and Jump Dynamics,” *Review of Financial Studies*, 21, 2535–2563.

——— (2012): “Jumps in Equilibrium Prices and Market Microstructure Noise,” *Journal of Econometrics*, 168, 396–406.

Mancini, C. (2001): “Disentangling the Jumps of the Diffusion in a Geometric Jumping Brownian Motion,” *Giornale dell’Istituto Italiano degli Attuari*, LXIV, 19–47.

Mykland, P. A., and L. Zhang (2006): “ANOVA for Diffusions and Itô Processes,” *Annals of Statistics*, 34, 1931–1963.

——— (2009): "Inference for continuous semimartingales observed at high frequency," *Econometrica*, 77, 1403–1455.

——— (2011): "The Double Gaussian Approximation for High Frequency Data," *Scandinavian Journal of Statistics*, 38, 215–236.

——— (2012): "The Econometrics of High Frequency Data," in *Statistical Methods for Stochastic Differential Equations*, ed. by M. Kessler, A. Lindner, and M. Sørensen, pp. 109–190. Chapman and Hall/CRC Press, New York.

Mykland, P. A., and L. Zhang (2016): "Between data cleaning and inference: Pre-averaging and other robust estimators of the efficient price," *Journal of Econometrics*, 194, 242–262.

Mykland, P. A., L. Zhang, and D. Chen (2019): "The algebra of two scales estimation, and the S-TSRV: High frequency estimation that is robust to sampling times," *Journal of Econometrics*, 208, 101–119.

Podolskij, M., and M. Vetter (2009a): "Bipower-type estimation in a noisy diffusion setting," *Stochastic Processes and Their Applications*, 119, 2803–2831.

——— (2009b): "Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps," *Bernoulli*, 15, 634–658.

Reiss, M. (2011): "Asymptotic equivalence for inference on the volatility from noisy observations," *Annals of Statistics*, 2, 772–802.

Reiss, M., V. Todorov, and G. Tauchen (2015): "Nonparametric test for a constant beta between Itô semi-martingales based on high-frequency data," *Stochastic Processes and their Applications*, 125, 2955–2988.

Renault, E., C. Sarisoy, and B. J. Werker (2017): "Efficient estimation of integrated volatility and related processes," *Econometric Theory*, 33, 439–478.

Shephard, N., and D. Xiu (2012): "Econometric analysis of multivariate realised QML: estimation of the covariation of equity prices under asynchronous trading," discussion paper.

Williams, D. (1991): *Probability with martingales*. Cambridge University Press.

Xiu, D. (2010): "Quasi-Maximum Likelihood Estimation of Volatility WIth High Frequency Data," *Journal of Econometrics*, 159, 235–250.

Zhang, L. (2001): "From martingales to ANOVA: implied and realized volatility," Ph.D. thesis, Department of Statistics, University of Chicago.

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

——— (2011): "Estimating Covariation: Epps Effect and Microstructure Noise," *Journal of Econometrics*, 160, 33–47.

——— (2012): "Implied and realized volatility: Empirical model selection," *Annals of Finance*, 8, 259–275.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005): "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," *Journal of the American Statistical Association*, 100, 1394–1411.