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

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.

*The authors thank Peter Phillips and Viktor Todorov, the Editors, as well as the Referees for insightful comments which substantially improved the paper. We thank Fangfang Wang for excellent research assistance. We also benefited from comments by participants at the University of Chicago Stevanovich Center for Financial Mathematics conference on volatility, the Imperial College financial econometrics conference, the North American Winter Meetings of the Econometric Society, Northern Illinois University, the Inaugural Conference of the Society for Financial Econometrics, the University of Pennsylvania, and the Canadian Econometric Study Group. We also thank Evan Anderson, Tim Bollerslev, Frank Diebold, Rob Engle, Nour Meddahi, Neil Shephard and Kevin Sheppard for comments. The authors are also grateful for support under NSF grants DMS 17-13129 and DMS 20-15544.

†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, 5747 S. Ellis Avenue, Chicago, Illinois 60637, e-mail: mykland@pascal.uchicago.edu.

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

1 Introduction

Substantial progress has been made on in-sample asymptotics, against the backdrop of increasingly available high frequency financial data. The asymptotic analysis pertains to statistics based on samples over finite intervals involving data observed at ever increasing frequency. The prime example is measures of increments in quadratic variation, see Jacod (1994), Jacod (1996) and Barndorff-Nielsen and Shephard (2002b) as well as the survey by Barndorff-Nielsen and Shephard (2007) and monographs by Jacod and Protter (2012), Mykland and Zhang (2012) and Aït-Sahalia and Jacod (2014).[1] The empirical measures attempt to capture volatility of financial markets, including possibly jumps. Moreover, a richly developed mathematical theory of semi-martingale stochastic processes provides the theoretical underpinning for measuring volatility in the context of arbitrage-free asset pricing models based on frictionless financial markets.

The aforementioned literature of measuring volatility has been the motivation for a now standard two-step modeling approach. The first step consists of measuring past realizations of volatility accurately over non-overlapping intervals - typically daily - and the second is to build models using the so called realized measures.

While persistence in volatility has been exploited extensively to predict future outcomes, it has not been exploited to improve upon the measurement of current and past realized volatility. It is shown in this paper that the in-sample asymptotics can be complemented with observations in prior intervals, that is in-sample statistics can benefit from across-sample observations.

While volatility is a lead example, our theory applies to many empirical processes that are important for financial analysis. First, we can easily extend our analysis to higher moments, beyond realized variances, such as kurtosis-related quarticity. The latter is used for feasible asymptotic distribution theory of high frequency data statistics. Since higher moments are known to be less precisely estimated, our analysis becomes even more relevant with finitely sampled data. Another prominent example is the estimation of realized betas. Our simulation results indeed indicate that substantial improvements can be made - more so than for realized volatility - by exploiting information about beta in prior intervals.

[1] Other examples include measure of bi-power and power variation as well as other functional transformations of returns sampled at high frequency (see again Barndorff-Nielsen and Shephard (2007), Jacod and Protter (2012), Mykland and Zhang (2012) or Aït-Sahalia and Jacod (2014)for relevant references).

We derive *conditional* filtering schemes, dependent on the path of the volatility process. Our filtering is therefore time-varying, meaning it is more efficient than unconditional filters, and most importantly cannot be by-passed or absorbed as part of a fixed parameter prediction model. Despite being conditional, our filtering scheme remains model-free and is based on prediction errors, rather than linear combinations of past and present realized volatilities. The model-free aspect is something our approach shares with Foster and Nelson (1996) and Andreou and Ghysels (2002). The analysis in this paper is quite similar in spirit to Kalman-Bucy filtering, with some important differences as we do not deal with a Gaussian system, yet to remain model-free, use linear projections. Our analysis is also in the spirit of ideas having to do with forecast combination for nested models, see in particular Clark and McCracken (2009).

The paper is organized as follows. We start in Section 2 by deriving all our results using a simple framework. The example is stylized for illustrative purpose - yet turns out to be surprisingly comprehensive. Section 3 covers the asymptotic theory for a general setting. Alternative weighting schemes are covered in Section 4. Section 5 reports simulation evidence on the efficiency and forecasting gains in a controlled environment of stylized diffusion processes and realized betas. Section 6 concludes the paper. Supplementary material is provided in the Online Appendix. The Online Appendix (a) covers the proof of Proposition 3.1, (b) furnishes additional volatility examples as well as examples beyond volatility, (c) provides insights on the connection of our analysis with the Kalman filter and (d) discusses a forecasting example.

2 Derivations with a Simplified yet Comprehensive Example

The purpose of this section is to start with a relatively simple example that contains the core ideas of our analysis. Hence, the example is stylized for illustrative purpose - yet as we will show later it turns out to be a surprisingly comprehensive example. We start with a time index t, which we think of as daily, or weekly, monthly etc. For simplicity we will assume a daily process, although the reader can keep in mind that *ceteris paribus* all the derivations apply to any level of aggregation. Henceforth, we will use "day" and "period" t interchangeably, although the former will only be used for convenience. Moreover, while

we consider exclusively equally spaced discrete sampling, one could also think of unequally spaced data.

Within every period t, we consider returns over short equal-length intervals (i.e. intra-daily). The return denoted as:

$$X_{t,j}^n = p_{t-(j-1)/n} - p_{t-j/n} \tag{2.1}$$

where $1/n$ is the (intra-daily) sampling frequency and $p_{t-(j-1)/n}$ is the log price of a financial asset at the end of the jth interval of day t, with $j = 1, \ldots, n$. For example, when dealing with typical stock market data we will use $n = 78$ corresponding to a five-minute sampling frequency. We start with the following assumption about the data generating process:

Assumption 2.1 *Within a day (period) t, given a sequence $\sigma_{t,j}^2$, $j = 1, \ldots, n$, the return process in equation (2.1) is distributed independently Gaussian for all $j = 1, \ldots, n$:*

$$X_{t,j}^n \sim \mathcal{N}\left(0, \frac{1}{n}\sigma_{t,j}^2\right). \tag{2.2}$$

For every period t the parameter of interest is:

$$\sigma_{n,t}^2 \equiv Var(\sum_{j=1}^{n}[X_{t,j}^n]) \equiv \frac{1}{n}\sum_{j=1}^{n}\sigma_{t,j}^2 \tag{2.3}$$

and consider the following ML estimators for each t :

$$\hat{\sigma}_{n,t}^2 = \sum_{j=1}^{n}[X_{t,j}^n]^2. \tag{2.4}$$

Then conditional on the volatility path $\sigma_{t,j}^2$, $j = 1, \ldots, n$, we have, under Assumption 2.1, the following properties for the ML estimators: $E_\sigma[\hat{\sigma}_{n,t}^2] = \sigma_{n,t}^2$, $Var_\sigma[\hat{\sigma}_{n,t}^2] = (2/n^2)\sum_{j=1}^{n}\sigma_{t,j}^4 = (2/n)\sigma_{n,t}^{[4]}$ where $\sigma_{n,t}^{[4]} = 1/n\sum_{j=1}^{n}\sigma_{t,j}^4$, $E_\sigma[\cdot] = E[\cdot|\sigma_{t,j}^2, \forall j]$ and similarly $Var_\sigma[\cdot] = Var[\cdot|\sigma_{t,j}^2, \forall j]$.

2.1 Optimal weighting scheme in finite samples

The ML estimator $\hat{\sigma}_{n,t}^2$ of $\sigma_{n,t}^2$ defined in equation (2.4) is obviously endowed with several optimal properties, both in finite samples and asymptotically when the number n of intraday data goes to infinity. However, for a given sample size n, we may revisit the intuition

put forward by Barndorff-Nielsen, Nielsen, Shephard, and Ysusi (2004) that more precise estimators could be obtained by pooling neighboring time series observations for realized variances.

It is worth keeping in mind that our goal is to improve upon a conditional MLE $\hat{\sigma}^2_{n,t}$ of $\sigma^2_{n,t}$ given the volatility path and, therefore, the pooling issue must also be addressed given the volatility path. In other words, even if we follow the linear filtering strategy of Barndorff-Nielsen, Nielsen, Shephard, and Ysusi (2004), we will be looking for an estimator $\bar{\sigma}^2_{n,t}$ of $\sigma^2_{n,t}$ that would be an affine function of past and current realized volatilities: $\bar{\sigma}^2_{n,t} = a_{n,t} + \sum_{h=0}^{H} b_{n,t-h}\hat{\sigma}^2_{n,t-h}$ where the coefficients $a_{n,t}$, $b_{n,t-h}$, $h = 0, 1, \ldots, H$ are possibly functions of the volatility path. In contrast, the filtering procedures put forward by Barndorff-Nielsen, Nielsen, Shephard, and Ysusi (2004) only involve constant coefficients. Of course, it would be an ill-posed problem to search for coefficients that would be arbitrary functions of the volatility path as an obvious choice would certainly be: $a_{n,t} = \sigma^2_t, b_{n,t-h} = 0, \forall\, h = 0, 1, \ldots, H$. The challenge is to look for functions $a_{n,t}$, $b_{n,t-h}$, $h = 0, 1, \ldots, H$ of the volatility path which would yield feasible estimators of $\sigma^2_{n,t}$ knowing that the various realizations of the volatility path, using their sample counterparts, leave some room for accuracy improvements with respect to the naive estimator $\hat{\sigma}^2_{n,t}$. In other words, some kind of shrinkage is called for with respect to the aforementioned ill-posed problem. In the spirit of forecast combinations, it is then natural to look for a convex combination of two possible "forecasts": $\hat{\sigma}^2_{n,t}(\omega_t) = (1-\omega_t)\hat{\sigma}^2_{n,t} + \omega_t\breve{\sigma}^2_{n,t|t-1}$ where only the weight ω_t would depend on the volatility path while the alternative forecast $\breve{\sigma}^2_{n,t|t-1}$ would be a feasible function of past observations (the use of the index $t|t-1$ will be explained shortly). The optimal weight $\omega^*_{n,t}$ should then be chosen such that it minimizes the conditional mean square error of prediction:

$$\begin{aligned} \omega^*_{n,t} &= Arg\min_{\omega_t} E_\sigma[\hat{\sigma}^2_t(\omega_t) - \sigma^2_{n,t}]^2 = Arg\min_{\omega_t} MSE_\sigma(\omega_t) \\ &= Arg\min_{\omega_t} E_\sigma\left\{\hat{\sigma}^2_{n,t} - \sigma^2_{n,t} - \omega_t(\hat{\sigma}^2_{n,t} - \breve{\sigma}^2_{n,t|t-1})\right\}^2. \end{aligned} \tag{2.5}$$

Note that since $\sigma^2_{n,t} = E_\sigma(\hat{\sigma}^2_{n,t})$, we do not want $\breve{\sigma}^2_{n,t|t-1}$ to be a forecast of $\sigma^2_{n,t-1}$ but of $\sigma^2_{n,t}$ which explains why we index it by $t|t-1$. To make it feasible, it is then natural to define $\breve{\sigma}^2_{n,t|t-1}$ as a forecast of $\hat{\sigma}^2_{n,t}$ based on past return observations $x_{\tau,j}$, $\tau < t, j = 1, \ldots, n$. Many forecasting strategies may be worth considering, including MIDAS regressions to use past high frequency information to "nowcast" the current realized volatility $\hat{\sigma}^2_{n,t}$. While these relevant extensions would not introduce any specific difficulty for practical implementation,

we will stick, just for the sake of notational simplicity, to a more conventional linear projection approach for the definition of $\breve{\sigma}^2_{n,t|t-1}$:

$$\breve{\sigma}^2_{n,t|t-1} = c_n + \sum_{h=1}^{H} \varphi_{n,h} \hat{\sigma}^2_{n,t-h} \tag{2.6}$$

where the coefficients c, φ_h, $h = 1, \ldots, H$ are defined as linear regression coefficients of $\hat{\sigma}^2_{n,t}$ on H lagged values:

$$\begin{aligned} \varphi_n^{(H)} &= (\varphi_{n,h})_{1\leq h\leq H} = \left(Var[(\hat{\sigma}^2_{n,t-h})_{1\leq h\leq H}]\right)^{-1} Cov\left[(\hat{\sigma}^2_{n,t-h})_{1\leq h\leq H}, \hat{\sigma}^2_{n,t}\right], \\ c_n &= \left(1 - \sum_{h=1}^{H} \varphi_{n,h}\right) E(\hat{\sigma}^2_{n,t}). \end{aligned} \tag{2.7}$$

Note that formulas (2.7) implicitly involve a stationarity assumption:

Assumption 2.2 *The n-dimensional process* $(\sigma^2_{t,j})_{1\leq j\leq n}, t \in \mathbb{Z}$, *is weakly stationary.*

The above Assumption implies that the estimated process $(\hat{\sigma}^2_{n,t}), t \in \mathbb{Z}$, is itself weakly stationary. In particular:

$$\begin{aligned} E(\hat{\sigma}^2_{n,t}) &= E(\sigma^2_{n,t}) = \sigma^2 \\ Cov\left[\hat{\sigma}^2_{n,t-h}, \hat{\sigma}^2_{n,t}\right] &= Cov[\sigma^2_{n,t-h}, \sigma^2_{n,t}] = \gamma_\sigma(h), h > 0 \\ Var(\hat{\sigma}^2_{n,t}) &= Var(\sigma^2_{n,t}) + \frac{2}{n} E\left(\sigma^{[4]}_{n,t}\right) = \gamma_\sigma(0) + \frac{2}{n}\nu_{\sigma,4} \end{aligned}$$

where $Var(\sigma^2_{n,t}) = \gamma_\sigma(0)$ and $E\left(\sigma^{[4]}_{n,t}\right) = \nu_{\sigma,4}$, so that formulas (2.7) can be rewritten:

$$\begin{aligned} \varphi_n^{(H)} &= (\varphi_{n,h})_{1\leq h\leq H} = \Sigma^{-1}_{n,H} \left[\gamma_\sigma(h)\right]_{1\leq h\leq H} \\ \Sigma_{n,H} &= Var[(\sigma^2_{n,t-h})_{1\leq h\leq H}] + \frac{2}{n}\nu_{\sigma,4} Id_H \\ c_n &= \left(1 - \sum_{h=1}^{H} \varphi_{n,h}\right) \sigma^2. \end{aligned}$$

Hence, we are interested in the minimization program (2.5) for the definition (2.6) of $\breve{\sigma}^2_{n,t|t-1}$.

A straightforward derivation yields:

$$\omega^*_{n,t} = \frac{Cov_\sigma\left[\hat{\sigma}^2_{n,t}, \hat{\sigma}^2_{n,t} - \breve{\sigma}^2_{n,t|t-1}\right]}{E_\sigma\left[\hat{\sigma}^2_{n,t} - \breve{\sigma}^2_{n,t|t-1}\right]^2} = \frac{Var_\sigma[\hat{\sigma}^2_{n,t}]}{E_\sigma\left[\hat{\sigma}^2_{n,t} - \breve{\sigma}^2_{n,t|t-1}\right]^2}. \tag{2.8}$$

To summarize, our preferred estimator is:

$$\hat{\sigma}^2_{n,t}(\omega^*_{n,t}) = (1 - \omega^*_{n,t})\hat{\sigma}^2_{n,t} + \omega^*_{n,t}\breve{\sigma}^2_{n,t|t-1}, \tag{2.9}$$

where: (i) $\breve{\sigma}^2_{n,t|t-1}$ is defined by regression on H lagged values: $\breve{\sigma}^2_{n,t|t-1} = c_n + \sum_{h=1}^{H} \varphi_{n,h}\hat{\sigma}^2_{n,t-h}$ and (ii) optimal weights $\omega^*_{n,t}$ are computed according to (2.8).

It is worth keeping in mind that, given the volatility path, the consecutive estimators $\hat{\sigma}^2_{n,t-h}$, $h = 0, 1, 2, \ldots$ are serially independent. This explains the second equality in (2.8) and will be repeatedly used in computations below. Note that $\omega^*_{n,t}$ is in general smaller than the conditional regression coefficient $b_{n,t}$ of $(\hat{\sigma}^2_{n,t} - \sigma^2_{n,t})$ on $(\hat{\sigma}^2_{n,t} - \breve{\sigma}^2_{n,t|t-1})$, due to the fact that the alternative estimator $\breve{\sigma}^2_{n,t|t-1}$ for $\sigma^2_{n,t}$ is biased:

$$\omega^*_{n,t} = \frac{Var_\sigma\left[\hat{\sigma}^2_{n,t}\right]}{Var_\sigma\left[\hat{\sigma}^2_{n,t} - \breve{\sigma}^2_{n,t|t-1}\right] + \left[E_\sigma(\breve{\sigma}^2_{n,t|t-1}) - \sigma^2_{n,t}\right]^2} \leq b_{n,t} = \frac{Var_\sigma\left[\hat{\sigma}^2_{n,t}\right]}{Var_\sigma\left[\hat{\sigma}^2_{n,t} - \breve{\sigma}^2_{n,t|t-1}\right]}.$$

This shrinkage of the coefficients of control variables when they are biased is well-known in the Monte Carlo literature (see e.g. Glynn and Iglehart (1989)) and easy to understand: (1) on the one hand, we want to replace our initial estimator $\hat{\sigma}^2_{n,t} = \hat{\sigma}^2_{n,t}(0)$ by $\hat{\sigma}^2_{n,t}(\omega_{n,t})$ for some positive weight $\omega_{n,t}$ because we think that some variance reduction is allowed by combining our initial estimator $\hat{\sigma}^2_{n,t}$ for $\sigma^2_{n,t}$ with another one $\breve{\sigma}^2_{n,t|t-1}$, while (2) on the other hand, the price to pay for this variance reduction is the introduction of some bias since our alternative estimator $\breve{\sigma}^2_{n,t|t-1}$ of $\sigma^2_{n,t}$ is biased: $E_\sigma(\breve{\sigma}^2_{n,t|t-1})$ - $\sigma^2_{n,t} = c_n + \sum_{h=1}^{H} \varphi_{n,h}\sigma^2_{n,t-h} - \sigma^2_{n,t}$ $= \sum_{h=0}^{H} \varphi_{n,h}(\sigma^2_{n,t-h} - \sigma^2)$, with $\varphi_{n,0} = -1$, so that:

$$E_\sigma(\hat{\sigma}^2_{n,t}(\omega_{n,t})) - \sigma^2_{n,t} = \omega_{n,t}\sum_{h=0}^{H} \varphi_{n,h}(\sigma^2_{n,t-h} - \sigma^2). \tag{2.10}$$

However, the optimal weight $\omega^*_{n,t}$ will become arbitrarily close to zero when the bias in the estimator $\breve{\sigma}^2_{n,t|t-1}$ is large. Hence, there is little cost to applying our optimal weighting

strategy since, if the bias is not as small as one may have hoped, the optimal weight brings us back to standard MLE. It is actually worth noting that optimal weight $\omega^*_{n,t}$ is a decreasing function of the variation coefficient $k_{n,t}$ of the estimation error $(\breve{\sigma}^2_{n,t|t-1} - \sigma^2_{n,t})$:

$$\begin{aligned} Var_\sigma\left[\hat{\sigma}^2_{n,t} - \breve{\sigma}^2_{n,t|t-1}\right] &= Var_\sigma\left[\hat{\sigma}^2_{n,t}\right] + Var_\sigma\left[\breve{\sigma}^2_{n,t|t-1}\right] = Var_\sigma\left[\hat{\sigma}^2_{n,t}\right] + \sum_{h=1}^{H} \varphi^2_{n,h} Var_\sigma\left[\hat{\sigma}^2_{n,t-h}\right] \\ &= \left[1 + \sum_{h=1}^{H} \psi^2_{n,h,t}\right] Var_\sigma\left[\hat{\sigma}^2_{n,t}\right], \end{aligned}$$

whereas: $\psi^2_{n,h,t} = \varphi^2_{n,h}[Var_\sigma[\hat{\sigma}^2_{n,t-h}]]/Var_\sigma[\hat{\sigma}^2_{n,t}]$. Letting $\psi^{(H)}_{n,t} = (\psi_{n,h,t})_{1\leq h\leq H}$:

$$\begin{aligned} \frac{1}{\omega^*_{n,t}} &= \frac{Var_\sigma\left[\hat{\sigma}^2_{n,t} - \breve{\sigma}^2_{n,t|t-1}\right] + \left[E_\sigma(\breve{\sigma}^2_{n,t|t-1}) - \sigma^2_{n,t}\right]^2}{Var_\sigma[\hat{\sigma}^2_{n,t}]} \\ &= 1 + \left\|\psi^{(H)}_{n,t}\right\|^2 + \frac{\left[E_\sigma(\breve{\sigma}^2_{n,t|t-1}) - \sigma^2_{n,t}\right]^2}{Var_\sigma[\hat{\sigma}^2_{n,t}]} \\ &= 1 + (1 + k^2_{n,t})\left\|\psi^{(H)}_{n,t}\right\|^2 \text{ where } k^2_{n,t} = \frac{\left[E_\sigma(\breve{\sigma}^2_{n,t|t-1}) - \sigma^2_{n,t}\right]^2}{\left[\left\|\psi^{(H)}_{n,t}\right\|^2 Var_\sigma[\hat{\sigma}^2_{n,t}]\right]} = \frac{\left[E_\sigma(\breve{\sigma}^2_{n,t|t-1} - \sigma^2_{n,t})\right]^2}{Var_\sigma\left(\breve{\sigma}^2_{n,t|t-1} - \sigma^2_{n,t}\right)}. \end{aligned}$$

Through the underlying forecasting problem, parsimony matters. Increasing the number H of lags in the regression (2.6) may allow to reduce the variation coefficient of the estimation error of our alternative estimator $\breve{\sigma}^2_{n,t|t-1}$ but will have a cost in terms of increasing the norm of the vector $\psi^{(H)}_{n,t}$ of (rescaled) regression coefficients. Note however that the initial regression coefficients $\varphi_{n,h}, h = 1, ..., H$ have been rescaled to take into account the fact that daily variance estimators may have different precisions. As always, the cost of lack of parsimony will be magnified when it comes to feasible forecasting, that is when substituting estimators into theoretical formulas. Note that the theoretical gain in conditional MSE provided by the replacement of $\hat{\sigma}^2_{n,t}$ by $\hat{\sigma}^2_{n,t}(\omega^*_{n,t})$ is given by: $G_\sigma(\omega^*_{n,t}) = Var_\sigma[\hat{\sigma}^2_{n,t}]$ - $E_\sigma[\hat{\sigma}^2_{n,t}(\omega^*_{n,t}) - \sigma^2_{n,t}]^2 = (\omega^*_{n,t})^2 E_\sigma\left[(\hat{\sigma}^2_{n,t} - \breve{\sigma}^2_{n,t|t-1})^2\right]$. If we have to use an estimator $\hat{\omega}^*_{n,t}$ of the theoretically optimal weight $\omega^*_{n,t}$ instead of the optimal one, we will modify the quadratic estimation error by an amount of: $[\hat{\sigma}^2_{n,t}(\hat{\omega}^*_{n,t}) - \sigma^2_{n,t}]^2$- $[\hat{\sigma}^2_{n,t}(\omega^*_{n,t}) - \sigma^2_{n,t}]^2 = (\omega^*_{n,t} - \hat{\omega}^*_{n,t})(\hat{\sigma}^2_{n,t}$ - $\breve{\sigma}^2_{n,t|t-1})\left(\hat{\sigma}^2_{n,t}(\hat{\omega}^*_{n,t}) + \hat{\sigma}^2_{n,t}(\omega^*_{n,t}) - 2\sigma^2_{n,t}\right)$. Roughly speaking, the cost of estimation error will not exceed the benefit in MSE as long as the estimation error $(\hat{\omega}^*_{n,t} - \omega^*_{n,t})$ on the optimal weight $\omega^*_{n,t}$ will not exceed the level of this weight. In other words, the improvement pro-

posed in estimation of $\sigma^2_{n,t}$ that we propose in this section will be real insofar as: (1) on the one hand, the optimal weight $\omega^*_{n,t}$ to assign to past information (as conveyed by $\breve{\sigma}^2_{n,t-1}$) is not too small, in particular due to the bias in $\breve{\sigma}^2_{n,t-1}$, and (2) on the other hand, the estimation error $(\hat{\omega}^*_{n,t} - \omega^*_{n,t})$ on this optimal weight is not too large.

It is of course an empirical issue that will be properly assessed by our simulation study in section 5. However, in the next subsection, we put forward a set of arguments, based on asymptotic theory, that may enhance the likelihood of the required property.

2.2 Asymptotic feasibility of the optimal weighting scheme

Following the discussion in the previous subsection, we want to study asymptotic conditions such that the optimal weight $\omega^*_{n,t}$ is not too small and the estimation error on this weight is not too large, in order to leave room for a possible improvement. We will denote by $\hat{\omega}^*_{n,t,T}, t = 1, 2, ...T$, the estimated weights obtained from a time series $\hat{\sigma}^2_{n,t}, t = 1 - H, ..., 1, 2, ...T$ of daily observations. We will consider that the above required conditions (optimal weight $\omega^*_{n,t}$ not too small and estimation error not too large) may be fulfilled when these quantities are bounded in probability. In other words, the purpose of this subsection is to provide primitive assumptions ensuring the conjunction of the two following conditions: (a) $1/\omega^*_{n,t}= O_P(1)$, and (b) $\left|\hat{\omega}^*_{n,t,T} - \omega^*_{n,t}\right| = O_P(1)$.

It is then worth rewriting the optimal weight as follows:

$$\begin{aligned} \frac{1}{\omega^*_{n,t}} &= 1 + \left\|\psi^{(H)}_{n,t}\right\|^2 + \frac{\left[E_\sigma(\breve{\sigma}^2_{n,t|t-1}) - \sigma^2_{n,t}\right]^2}{Var_\sigma[\hat{\sigma}^2_{n,t}]} \\ &= 1 + \left\|\psi^{(H)}_{n,t}\right\|^2 + \frac{nB^2_{I,n}(t)}{2\sigma^{[4]}_{n,t}} + \frac{n\left[B^2_{F,n}(t) - B^2_{I,n}(t)\right]}{2\sigma^{[4]}_{n,t}}, \end{aligned} \tag{2.11}$$

where $B_{F,n}$ and B_I, n stand respectively for feasible and infeasible bias:

$$\begin{aligned} B_{F,n}(t) &= \sigma^2_{n,t} - \sum_{h=1}^{H} \varphi_{n,h}\sigma^2_{n,t-h} - \left(1 - \sum_{h=1}^{H} \varphi_{n,h}\right)\sigma^2 \\ \varphi^{(H)}_n &= (\varphi_{n,h})_{1\le h\le H} = \left(Var[(\hat{\sigma}^2_{n,t-h})_{1\le h\le H}]\right)^{-1} Cov\left[(\hat{\sigma}^2_{n,t-h})_{1\le h\le H}, \hat{\sigma}^2_{n,t}\right] \end{aligned}$$

while:

$$\begin{aligned} B_{I,n}(t) &= \sigma^2_{n,t} - \sum_{h=1}^{H} \varphi^0_{n,h}\sigma^2_{n,t-h} - \left(1 - \sum_{h=1}^{H} \varphi^0_{n,h}\right)\sigma^2 \\ \varphi^{0,(H)}_n &= (\varphi^0_{n,h})_{1\le h\le H} = \left(Var[(\sigma^2_{n,t-h})_{1\le h\le H}]\right)^{-1} Cov\left[(\sigma^2_{n,t-h})_{1\le h\le H}, \sigma^2_{n,t}\right]. \quad (2.12) \end{aligned}$$

Note that we use the word "feasible" with some abuse of language since the formula above for $B_{F,n}(t)$ still entails some unknown parameters that will have to be estimated. However, we will note later on that this estimation part is straightforward. Moreover, standard regularity conditions will obviously ensure that: $\left\|\psi^{(H)}_{n,t}\right\|^2 = O_P(1)$.

Then, we deduce from the decomposition (2.11) that a pair of sufficient conditions for the required property $1/\omega^*_{n,t} = O_P(1)$ is given by the following. First,

$$B_{I,n}(t) - B_{F,n}(t) = -\sum_{h=1}^{H}(\varphi^0_{n,h} - \varphi_{n,h})(\sigma^2_{n,t-h} - \sigma^2) = O_P(1/\sqrt{n}). \quad (2.13)$$

Second:

$$B_{I,n}(t) = O_P(1/\sqrt{n}). \quad (2.14)$$

Note that the conjunction of these two conditions implies in particular that:

$$B^2_{F,n}(t) - B^2_{I,n}(t) = (B_{F,n}(t) - B_{I,n}(t))\,(2B_{I,n}(t) + B_{F,n}(t) - B_{I,n}(t)) = O_P(1/n).$$

We know that by definition: $\Sigma_{n,H}\varphi^{(H)}_n = [\gamma_\sigma(h)]_{1\le h\le H}$, $\Sigma_{n,H} = \Gamma(H) + \frac{2}{n}\nu_{\sigma,4}Id_H, \Gamma(H) = \left[Var(\sigma^2_{n,t-h})_{1\le h\le H}\right]$, and $\Gamma(H)\varphi^{0,(H)}_n = [\gamma_\sigma(h)]_{1\le h\le H}$. Similarly, for $\varphi^{0,(H)}_n = (\varphi^0_{n,h})_{1\le h\le H}$. Therefore, by difference:

$$\Gamma(H)\left[\varphi^{(H)}_n - \varphi^{0,(H)}_n\right] + \frac{2}{n}\nu_{\sigma,4}\varphi^{(H)}_n = 0$$

and hence $\varphi^{(H)}_n$ - $\varphi^{0,(H)}_n = (-2/n)\nu_{\sigma,4}[\Gamma(H)]^{-1}\varphi^{(H)}_n = O(1/n)$ and by (2.13) the required bound follows:$|B_{I,n}(t) - B_{F,n}(t)| \le \|\varphi^{(H)}_n - \varphi^{0,(H)}_n\| \, \|\left(\sigma^2_{n,t-h} - \sigma^2\right)_{1\le h\le H}\| = O_P(1/n)$.

Regarding the second condition (2.14), let us illustrate it in the case $H = 1 : \sigma^2_{n,t} = (1 - \varphi^0_n)\,\sigma^2$ $+ \varphi^0_n\sigma^2_{n,t-1} + B_{I,n}(t)$. By definition: $E[(B_{I,n}(t))^2] = Var(B_{I,n}(t)) = [1 - (\varphi^0_n)^2]Var[\sigma^2_{n,t}]$. Therefore, we will ensure that $B_{I,n} = O_P(1/\sqrt{n})$ if we assume that:

Assumption 2.3 *Let $H = 1$ and φ_n^0, defined in (2.12) satisfy:*

$$1 - (\varphi_n^0)^2 = O(1/n). \tag{2.15}$$

Assumption 2.3 should be interpreted as a sufficient condition for $B_I = O_p(1/\sqrt{n})$. According to (2.11), a violation of this condition may imply that on some specific days, $1/\omega^*$ is not upper bounded, so that it will sometimes be optimal to overlook information about volatility from previous days. There is widely documented evidence of the existence of days with jumps in price and/or in volatility. Obviously, for these days, Assumption 2.3 is not realistic and pooling information about the volatility process across past days might be inappropriate. The logic for our belief that Assumption 2.3 should be relevant for most of days rests upon the new framework of near integration as formulated by Phillips, Moon, and Xiao (2001) (PMX hereafter) and described in the next subsection. Note that a similar conclusion would be easily obtained for $H > 1$ at the cost of more complicated notation.

We would like to emphasize that volatility jumps may come in the form of many small jumps, possibly of infinite activity. See, in particular, Jacod and Todorov (2010), Todorov and Tauchen (2011), Jacod and Protter (2012), and Aït-Sahalia and Jacod (2014). This situation represents an intermediate scenario between a single jump and a continuous volatility process. As shown in Zhang (2007), there may be contiguity between the probability distribution of a continuous (in this case volatility) process and one that has infinite activity jumps, thus strengthening the idea that such frequent jumps may in many cases be more like continuous evolution that the situation of a single free-standing jump. A proper characterization of this problem would be of substantial interest but is beyond the scope of this paper.

Regarding the estimation error, we have: $|\hat{\omega}^*_{n,t,T} - \omega^*_{n,t}| = (\omega^*_{n,t})(\hat{\omega}^*_{n,t,T})|(\omega^*_{n,t})^{-1} - (\hat{\omega}^*_{n,t,T})^{-1}| \leq |(\omega^*_{n,t})^{-1} - (\hat{\omega}^*_{n,t,T})^{-1}|$. Since: $(\omega^*_{n,t})^{-1} = 1 + \|\psi^{(H)}_{n,t}\|^2 + [E_\sigma(\breve{\sigma}^2_{n,t|t-1}) - \sigma^2_{n,t}]^2/Var_\sigma[\hat{\sigma}^2_{n,t}] = 1 + \|\psi^{(H)}_{n,t}\|^2 + [\sum_{h=0}^H \varphi_{n,h}(\sigma^2_{n,t-h} - \sigma^2)]^2/[(2/n)\sigma^{[4]}_{n,t}]$, we will define an estimator as:

$$\frac{1}{\hat{\omega}^*_{n,t,T}} = 1 + \left\|\hat{\psi}^{(H)}_{n,t,T}\right\|^2 + \frac{\left[\hat{c}_{n,T} + \sum_{h=1}^H \hat{\varphi}_{n,h,T}\hat{\sigma}^2_{n,t-h} - \hat{\sigma}^2_{n,t}\right]^2}{\frac{2}{n}\hat{\sigma}^{[4]}_{n,t}} = 1 + \left\|\hat{\psi}^{(H)}_{n,t,T}\right\|^2 + \frac{n}{2}\frac{\hat{\nu}^2_{n,t,T}}{\hat{\sigma}^{[4]}_{n,t}}, \tag{2.16}$$

where, from a time series $(\hat{\sigma}^2_{n,t})_{1-H\leq t\leq T}$, we can compute the OLS estimator $(\hat{c}_{n,T}, \hat{\varphi}^{(H)}_{n,T})$ in

the regression equation:

$$\hat{\sigma}_{n,t}^2 = c_n + \sum_{h=1}^{H} \varphi_{n,h} \hat{\sigma}_{n,t-h}^2 + v_{n,t}, t = 1, 2, ..., T$$

and $\hat{\nu}_{n,t,T} = \hat{\sigma}_{n,t}^2$ - $\hat{c}_{n,T} - \sum_{h=1}^{H} \hat{\varphi}_{n,h,T} \hat{\sigma}_{n,t-h}^2$ stands for the residual of this regression. Note that we never assume in this section that the processes σ_t^2 and $\hat{\sigma}_t^2$ are $AR(H)$ processes. The error term $v_{n,t}$ is not assumed to be a white noise, but only the residual of a regression of $\hat{\sigma}_{n,t}^2$ on H lagged values. We will discuss later the way to efficiently compute an estimator $\hat{\sigma}_{n,t}^{[4]}$ of $\sigma_t^{[4]}$, but obviously, the corresponding estimation error will be $O_P(1/\sqrt{n})$. Note that this estimation strategy will also be relevant to compute the coefficients of the vector $\hat{\psi}_{n,t,T}^{(H)}, h = 1, ..., H$: $\hat{\psi}_{n,h,t,T}^2 = (\hat{\sigma}_{n,t-h}^{[4]} / \hat{\sigma}_{n,t}^{[4]}) \hat{\varphi}_{n,h,T}^2$.

The key issue is now to assess the long-range asymptotics estimation error $\left|\hat{\omega}_{n,t,T}^* - \omega_{n,t}^*\right|$ as a function of the sample size T. By an argument similar to the one used to get (2.11), we can write with obvious notations: $(\hat{\omega}_{t,T}^*)^{-1} = 1 + \|\hat{\psi}_{n,t,T}^{(H)}\|^2 + n\hat{B}_{I,n,T}^2(t)/(2\hat{\sigma}_{n,t}^{[4]})$ and therefore: $(\hat{\omega}_{n,t,,T}^*)^{-1}$ -$(\omega_{n,t}^*)^{-1} = \|\hat{\psi}_{n,t,T}^{(H)}\|^2$ - $\|\psi_{n,t}^{(H)}\|^2 + n\hat{B}_{I,n,T}^2(t)/(2\hat{\sigma}_{n,t}^{[4]})$ - $nB_{I,n}^2(t)/(2\sigma_{n,t}^{[4]})$ -$n\left[B_{F,n}^2(t) - B_{I,n}^2(t)\right]/(2\sigma_{n,t}^{[4]})$.

From long range asymptotics: $|\|\hat{\psi}_{n,t,T}^{(H)}\|^2$ - $\|\psi_{n,t}^{(H)}\|^2| = O_P(1/T) + O_P(1/n)$. In order to control the difference $\left[n\hat{B}_{I,n,T}^2(t) - nB_{I,n}^2(t)\right]$, we will refer again to the near integration argument (as we did for condition (2.14)), with the illustrative case $H = 1$: $\hat{B}_{I,n,T}(t) = \hat{\sigma}_{n,t}^2$ - $\hat{c}_{n,T}$ - $\hat{\varphi}_{n,T}\hat{\sigma}_{n,t-1}^2$, so that:$|\hat{B}_{I,n,T}(t) - B_{I,n}(t)| = O_P(1/\sqrt{T}) + O_P(1/\sqrt{n})$. Moreover: $|\hat{B}_{I,n,T}(t) + B_{I,n}(t)| = |2B_{I,n}(t) + \hat{B}_{I,n,T}(t) - B_{I,n}(t)| \leq 2|B_{I,n}(t)| + |\hat{B}_{I,n,T}(t) - B_{I,n}(t)| = O_P(1/\sqrt{T})$ $+ O_P(1/\sqrt{n})$. Hence: $n|\hat{B}_{I,n,T}^2(t) - B_{I,n}^2(t)| = n|\hat{B}_{I,n,T}(t) - B_{I,n}(t)||\hat{B}_{I,n,T}(t) + B_{I,n}(t)| = O_P(1) + O_P(n/T) + O_P\left(\sqrt{n}/\sqrt{T}\right)$. Therefore, if we assume $n = O(T)$, we will obtain the required condition:

$$n\left|\hat{B}_{I,n,T}^2(t) - B_{I,n}^2(t)\right| = O_P(1).$$

In other words, if one uses five-minute returns in a 24-hour financial market ($n = 288$), we assume that more than 288 observation days are available.

2.3 A block local to unity framework for volatility

For the sake of expositional simplicity, we focus in this subsection on the case where only $H = 1$ lag is used to compute the forecast $\breve{\sigma}^2_{n,t-1}$ and the weighted estimator. The goal is to show that a natural block-local-to-unity framework for volatility in the spirit of PMX allows us to see the daily volatility persistence parameter φ^0_n as a function of n, such that the wished condition (2.15) is fulfilled:

$$n[1 - (\varphi^0_n)^2] = O(1). \tag{2.17}$$

To get the main intuition, let us imagine that within each "block" or "day" t, the volatility process $\sigma^2_{t,j}, j = 1, ..., n$, can be embedded into a continuous time diffusion process with linear drift:

$$d\sigma^2_t = \kappa \left[\sigma^2 - \sigma^2_t\right] dt + \gamma_t dW_t, k > 0$$

where W_t is a Wiener process and the length of a block (a day) is one unit of time. Note that an Ornstein-Uhlenbeck-like process as put forward by Barndorff-Nielsen and Shephard (2001) would also do the job. Indeed, our argument of embedding a discrete time volatility process of interest into a continuous time one is only for the purpose of studying second order moments when the sampling frequency n goes to infinity. It does not depend on higher order characteristics of the underlying stationary continuous time process. In particular we can apply formulas (44) and (45) of Barndorff-Nielsen and Shephard (2001) to conclude that the correlation coefficient between $\int_{t-1}^{t} \sigma^2(u)du$ and $\int_{t-2}^{t-1} \sigma^2(u)du$ is given by: $.5\left[1 - \exp(-\kappa)\right]^2 / \left[\exp(-\kappa) - 1 + \kappa\right]$. Note that by a Riemann integration argument: $\sigma^2_{n,t} = \frac{1}{n}\sum_{j=1}^{n} \sigma^2_{t,j} \underset{n\to\infty}{\longrightarrow} \int_{t-1}^{t} \sigma^2(u)du$. Therefore, under standard regularity conditions, we have for large n (see e.g. Andersen, Bollerslev, and Meddahi (2004)): $\varphi^0_n = Corr\left[\sigma^2_{n,t}, \sigma^2_{n,t-1}\right] \approx 0.5$ $[1 - \exp(-\kappa)]^2/[\exp(-\kappa) - 1 + \kappa]$. The idea of block local to unity is that within each day, volatility is highly persistent so that the mean reversion parameter κ is close to zero that the n daily observations can hardly detect its discrepancy from zero: $\kappa \approx c/n$, $c > 0$. In other words, the correlation coefficient between two values of the (squared) volatility process separated by Δ units of time is equivalent to $\exp(-c\Delta/n) \approx 1$ - $(c\Delta/n)$. Then, the correlation between two consecutive daily integrated variances will be: $0.5\left[1-\exp(-\kappa)\right]^2/[\exp(-\kappa)-1+\kappa]$ $\approx \left[1 - \frac{\kappa}{2}\right]^2 \approx 1 - \frac{c}{n}$. Then we obtain the result put forward in equation (2.17): $\varphi^0_n \approx 1 - \frac{c}{n}$ $\Rightarrow 1 - (\varphi^0_n)^2 \approx 1 - \left[1 - \frac{c}{n}\right]^2 \approx 2\frac{c}{n} \Rightarrow n\left[1 - (\varphi^0_n)^2\right] = O(1)$. Therefore, our maintained assumption appearing in equation (2.15) is implied by the block-local-to-unity assumption of

PMX. Note however an important difference between the original local-to-unity asymptotics and our use of it which is conformable to PMX. While the former near-to-unit root literature (see Bobkoski (1983), Chan and Wei (1987), Phillips (1987), Elliott, Rothenberg, and Stock (1996), among others) focuses on persistence parameters going to one at rate $1/T$, where T is the length of the time series, the rate of convergence in (2.17) is governed by n, i.e. the number of intradaily data. In this respect, what is really required for our approach is in fact:

$$[1-(\varphi_n^0)^2] = O(\varphi_n^0 - \varphi_n) \tag{2.18}$$

where the notation $O(.)$ must be understood as an upper bound. Note that $[1-(\varphi_n^0)^2]$ and $(\varphi_n^0 - \varphi_n)$ are two different objects and there is no obvious reason why they would converge at the same rate. In the sequel, the rate of convergence of $(\varphi_n^0 - \varphi_n)$ will sometimes be slower than $1/n$. It will notably depend on the quality of the volatility process estimator which may for example be corrupted by exogenous phenomena such as microstructure noise. The key assumption driving equation (2.18) is that, roughly speaking, the level of volatility persistence is at least as good as the quality of our intradaily volatility estimator. It ensures that the squared infeasible bias:

$$B_{I,n}^2(t) = O([1-(\varphi_n^0)^2]) = O(\varphi_n^0 - \varphi_n) \tag{2.19}$$

does not dominate the conditional variance $Var_\sigma(\hat{\sigma}_{n,t}^2 - \breve{\sigma}_{n,t|t-1}^2)$. Note also that for given n, the time series $(\sigma_{n,t}^2)$ and $(\hat{\sigma}_{n,t}^2)$ are still stationary processes, albeit highly persistent when n is large. In particular, a long time series $(\hat{\sigma}_{n,t}^2), t = 1, ..., T$ allows to consistently estimate (at standard rate $\sqrt{T}$) the unconditional mean σ^2 and the correlation coefficient φ_n. We will always consider that the time span is large enough to make $(1/T)$ small compared to $(1/n)$ (or at least $O(\varphi_n^0 - \varphi_n)$) in order to make negligible the time series estimation error in the parameters σ^2 and φ_n. For all practical purpose, our Monte Carlo experiments have confirmed that the time series estimation error on the parameters σ^2 and φ_n is negligible in front of $O(\varphi_n^0 - \varphi_n)$.

2.4 Estimating Optimal Weights

The optimal weighting scheme $\omega_{n,t}^*$ depends on $\sigma_{n,t}^{[4]}$ and φ_n. Because of the maintained practical assumption discussed above, we do not need to worry about the latter, hence our

focus will be on the former. The sample counterpart of $\sigma^{[4]}_{n,t}$, the so-called realized quarticity (see below), is known to be a very noisy estimator and therefore, albeit root-n consistent, may be detrimental regarding our search for efficiency gain.

Following a suggestion by Mykland and Zhang (2009), we rather consider the MLE estimation of $Var\hat{\sigma}^2_{n,t} = 2\sigma^{[4]}_{n,t}/n$ where: $\sigma^{[4]}_{n,t} = \frac{1}{n}\sum_{j=1}^{n}\sigma^4_{t,j}$, and efficiency gains are made possible by the following assumption:

Assumption 2.4 *Assume that n is a multiple of $m \geq 1$, and for $(i-1)m < j \leq im$, we have:*

$$\sigma_{t_j} = \sigma_{t_i} \qquad i = 1, \ldots, n/m.$$

Given Assumption 2.4 the MLE $\hat{\sigma}^2_{n,t,[i]}$ of $\sigma^2_{t_i}$is: $\hat{\sigma}_{n,t,[i]}= \frac{n}{m}\sum_{j=m(i-1)+1}^{mi} X^2_{t,j}$. Then the MLE of $\sigma^4_{n,t,[i]}$ is such that:

$$\frac{\hat{\sigma}^4_{n,t,[i]}}{\sigma^4_{n,t,[i]}} = \frac{1}{m^2}\frac{[n\sum_{j=m(i-1)+1}^{mi} X^2_{t,j}]^2}{\sigma^4_{n,t,[i]}} \backsim \frac{[\chi^2(m)]^2}{m^2}$$

with expectation $(1+2/m)$. Hence, an unbiased estimator of $\sigma^{[4]}_{n,t}$ is defined as:

$$\hat{\sigma}^{[4]}_{n,t} = \frac{m}{n}\sum_{i=1}^{n/m}\frac{\hat{\sigma}^4_{n,t,[i]}}{1+2/m} = \frac{n}{m+2}\sum_{i=1}^{n/m}\left[\sum_{j=m(i-1)+1}^{mi} X^2_{t,j}\right]^2 \tag{2.20}$$

whereas an estimator not taking advantage of $m > 1$ would be the realized quarticity:

$$\tilde{\sigma}^{[4]}_{n,t} = \frac{n}{3}\sum_{j=1}^{n} X^4_{t,j} = \frac{n}{3}\sum_i \sigma^4_{n,t,[i]}\sum_j (\frac{X_{t,j}}{\sigma_{n,t,[i]}})^4 \backsim \frac{1}{3n}\sum_i \sigma^4_{n,t,[i]}\sum_j (\chi^2(1))^2. \tag{2.21}$$

In Appendix A we compare the efficiency of the estimators $\hat{\sigma}^4_{n,t}$ and $\tilde{\sigma}^4_{n,t}$, showing that when $m > 1$, the former will be more efficient. This efficiency gain by grouping has a much more general validity (see Mykland and Zhang (2009)) and has been shown to be arbitrarily close to semiparametric efficiency in a very general setting (see Renault, Sarisoy, and Werker (2017)).

Finally, let us recall that we try to improve the estimation of $\sigma^2_{n,t}$ using prior day information, and in particular using $\hat{\sigma}^2_{n,t-1}$. This argument is not confined to volatility measures. In particular, we can use the arguments spelled out so far to improve upon $\hat{\sigma}^{[4]}_{n,t}$ by using estimates

from prior observation intervals. Namely, we can also consider the correlation coefficient ψ_n of $\hat{\sigma}^{[4]}_{n,t}$ on $\hat{\sigma}^{[4]}_{n,t-1}$:

$$\psi_n = \frac{Cov(\hat{\sigma}^{[4]}_{n,t}, \hat{\sigma}^{[4]}_{n,t-1})}{Var(\hat{\sigma}^{[4]}_{n,t})} \tag{2.22}$$

and develop a filtering procedure for $\sigma^{[4]}_{n,t}$ similar to the one proposed above for $\sigma^2_{n,t}$.

Finally, it should also be noted that our way to take advantage of past information is quite similar in spirit to the Kalman-Bucy filter but with a couple of important differences which are explained in Online Appendix Section OA.4.

3 Asymptotic Analysis for General Estimators

The example in the previous section - where we started with a relatively simple case of a piecewise constant volatility process - is surprisingly comprehensive. The purpose of the present section is to aim for generality.

3.1 The case of no leverage and drift

We start by replacing the specification of the data generating process in Assumption 2.1 with:

Assumption 3.1 *We suppose that X_s is an Itô-semimartingale, either without jumps*

$$dX_s = \mu(s)ds + \sigma(s)dW(s),$$

or with jumps that are removed (see Footnote 2). We also assume for the moment - but relax later - that $\mu(s) \equiv 0$ and that the volatility $\sigma(s)$, is independent of the Brownian motion $W(s)$.

In the case of jump diffusions, we will leave unspecified how a researcher goes about removing jumps.[2] Next, we consider the generic setting examined by Mykland and Zhang (2017),

[2] Such as, for example, by (1) using bi- and multi-power methods (Barndorff-Nielsen and Shephard (2004b, 2006), Mykland, Shephard, and Sheppard (2012), or (2) by truncation (Mancini (2001), Aït-Sahalia

namely one seeks to estimate:

$$\Theta_t = \int_{t-1}^{t} \theta(s)ds, \tag{3.1}$$

where $\{\theta(s)\}$ is a spot parameter process, such as squared spot volatility, leverage, an instantaneous regression parameter, among others.

Similar to the analysis in the previous section, we have n discretely sampled data points which yield an estimator $\hat{\Theta}_{n,t}$, for which we have the following properties:

Assumption 3.2 *As $n \to \infty$:*

(i) The estimator $\hat{\Theta}_{n,t}$ is consistent

(ii) A CLT applies, namely $n^{\alpha}(\hat{\Theta}_{n,t} - \Theta_t) \xrightarrow{\mathcal{L}} V_t^{1/2} N(0,1)$ in law, conditionally on the process $\sigma(s)$, and V_t is a measurable function of the process $(\theta(s))_{t-1<s\leq t}$.

In the remainder of this section we consider various estimators $\hat{\Theta}_{n,t}$ proposed in the literature, which we intend to generalize to estimators of the type appearing in equation (2.9), denoted $\hat{\Theta}_{n,t}(\omega_t^*)$, using the arguments advanced in the previous section. In addition to Assumptions 3.1 and 3.2, we also need an assumption akin to the block local to unity framework discussed in section 2.3. Because the generic setting for Θ_t covers a wide variety of cases, we cannot state this as a generic assumption. For example, in the case of volatility related applications, it will require that the volatility process is sufficiently persistent (not specified in Assumption 3.1). In other applications, such as leverage, the persistence actually applies to $\mu(s)$ in Assumption 3.1.[3]

Throughout, we study the case of equidistant sampling, $t_i - t_{i-1} = \Delta\ t_i = 1/n$. We may observe X_{t_i} at times $t_i,\ i = 0, \cdots, n$ spanning $(t-1, t]$.[4] Alternatively, if there is microstructure noise, the observations are of the form

$$Y_{t_i} = X_{t_i} + \epsilon_{t_i}, \tag{3.2}$$

and Jacod (2007, 2008, 2009, 2012), Jacod and Todorov (2010), Lee and Mykland (2008, 2012), Jing, Kong, Liu, and Mykland (2012)), or (3) with the help of the empirical characteristic functions (Jacod and Todorov (2014, 2018)).

[3]For specific cases we will discuss the equivalence of equation (2.18). See for example equations (3.9) (3.24) and later in this section.

[4]Note that henceforth we use notation slightly different from equation (2.1) and the previous section - as it is more convenient to do so henceforth.

where the noise is either i.i.d., or is stationary with fast mixing dependence.

To provide a few initial examples of estimators satisfying Conditions [i], and [ii], consider the following (non-exhaustive list of) volatility estimators. Here, $\theta(s) = \sigma^2(s)$.

Example 1 (Realized Volatility, No Microstructure Noise) The convergence rate is $\alpha = 1/2$. If X is continuous case, the baseline estimator for the $\int_{t-1}^{t} \theta(s)ds$ is the standard realized volatility (RV), $\sum_{t-1<t_{i+1}\leq t}(X_{t_{i+1}} - X_{t_i})^2$ (Andersen, Bollerslev, Diebold, and Ebens (2001), Andersen, Bollerslev, Diebold, and Labys (2001), Barndorff-Nielsen and Shephard (2002a)). The consistency and stable convergence has been shown by Jacod and Protter (1998) using discretization. □

Example 2 (Bipower Variation, No Microstructure Noise) The bipower variation $\hat{\Theta}_{n,t} = \frac{\pi}{2}\sum_{t-1<t_i\leq t} |\Delta X_{t_{i-1}}||\Delta X_{t_i}|$ (and more generally, multipower variation, Barndorff-Nielsen and Shephard (2004b), Barndorff-Nielsen and Shephard (2006)) estimates the integrated volatility in a way that is robust to jumps (where we specify that $dX_s = \mu(s)ds + \sigma(s)dW(s) + dJ(s)$, where $J(s)$ is a pure jump semimartingale). The convergence rate is $\alpha = 1/2$. Consistency and stable convergence has been shown in the papers by Barndorff-Nielsen and Shephard (2004b), Barndorff-Nielsen and Shephard (2006), Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006a), and Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006b). □

Starting with Example 1, 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. $\lim_{n\to\infty} \hat{\Theta}_{n,t} \to^p \int_{t-1}^{t} \theta(s)ds$. The criterion to minimize will be the conditional mean squared error:

$$E_\sigma \left[\hat{\Theta}_{n,t}(\omega_t) - \Theta_t\right]^2 = E_\sigma \left\{\hat{\Theta}_{n,t} - \Theta_t - \omega_t \left(\hat{\Theta}_{n,t} - \breve{\Theta}_{n,t|t-1}\right)\right\}^2. \tag{3.3}$$

Then, the problem to solve is obviously nearly identical to the one considered in Section 2, so that:

$$\hat{\Theta}_{n,t}(\omega^*_{n,t}) = \hat{\Theta}_{n,t} - \omega^*_{n,t}\left(\hat{\Theta}_{n,t} - \breve{\Theta}_{n,t|t-1}\right) \tag{3.4}$$

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

variable formula:

$$\omega^*_{n,t} = \frac{Cov_\Theta\left[\hat{\Theta}_{n,t}, \hat{\Theta}_{n,t} - \breve{\Theta}_{n,t|t-1}\right]}{E_\Theta\left[\hat{\Theta}_{n,t} - \breve{\Theta}_{n,t|t-1}\right]^2} = \frac{Var_\Theta[\hat{\Theta}_{n,t}]}{E_\Theta\left[\hat{\Theta}_{n,t} - \breve{\Theta}_{n,t|t-1}\right]^2}. \tag{3.5}$$

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.

Proposition 3.1 *For Example 1, the standard realized volatility (RV), obtained by summing squared intra-daily returns, yielding the so called realized variance, namely:*

$$\hat{\Theta}_{n,t} = \sum_{t-1<t_{i+1}\leq t} (X_{t_{i+1}} - X_{t_i})^2. \tag{3.6}$$

Then, under Assumptions 3.1 and 3.2, in analogy with equation (2.11), the optimal weight can be written as:

$$\begin{aligned} \frac{1}{\omega^*_{n,t}} &= 1 + \varphi_n^2 \frac{V_{t-1}}{V_t} + n\frac{B^2_{F,n}(t) - B^2_I(t)}{V_t} + n\frac{B^2_I(t)}{V_t} + o\left(\frac{1}{n}\right) \\ B_I(t) &= \Theta_t - \varphi^0\Theta_{t-1} - (1-\varphi^0)E\left[\Theta_t\right] \\ \varphi^0 &= \frac{Cov\left[\Theta_t, \Theta_{t-1}\right]}{Var\left(\Theta_t\right)} = \varphi_n + O\left(\frac{1}{\sqrt{n}}\right) \end{aligned} \tag{3.7}$$

where $V_t = 2\int_{t-1}^t \sigma(s)^4 ds$ *is called (twice) the integrated quarticity and* $B_{F,n}(t) = \Theta_t$ - $\varphi_n\Theta_{t-1} - (1-\varphi_n)E\left[\Theta_t\right]$ *with* $\varphi_n = Cov(\hat{\Theta}_{n,t}, \hat{\Theta}_{n,t-1})/Var(\hat{\Theta}_{n,t-1})$, *in the case of an AR(1) prediction model. For Example 2, using bi-power variation (BPV) defined as:*

$$\hat{\Theta}_{n,t}(k) = \frac{\pi}{2}\sum_{j=k+1}^{n} \left|X_{n,t_j}\right| \left|X_{n,t_{j-k}}\right|, \tag{3.8}$$

without loss of generality, setting $k = 1$, *it will be an optimal improvement of* $\hat{\Theta}_{n,t}$ *when* ω^*_t *is again defined according to the control variable formula (3.5) 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.*

Given the similarity with the analysis in the previous section, we skip the details here, as they appear in the Online Appendix. In order to estimate volatility on day t, we attach a non-zero weight ω_t^* to volatility information on day $t-1$. This weight increases as the relative size of the asymptotic variance V_t/n of $\hat{\Theta}_{n,t}$ is large in comparison to both (1) the asymptotic variance V_{t-1}/n of $\hat{\Theta}_{n,t-1}$ as well as (2) the squared forecast bias $B_{F,n}^2(t)$. However, for a given non-zero asymptotic bias $B_I(t)$, the term $n\ B_I^2(t)$ goes to infinity when n goes to infinity, likely pushing to zero the optimal weight $\omega_{n,t}^*$. The reason is fairly straightforward: since $\hat{\Theta}_{n,t}$ is a consistent estimator of $\int_{t-1}^t \theta(s)ds$, forecasting $\int_{t-1}^t \theta(s)ds$ from Θ_{t-1} becomes irrelevant when n becomes infinitely large: even a small forecast error has more weight than a vanishing estimation error. However, in practice, n is never infinitely large and there likely is a sensible trade-off between estimation error as measured by the asymptotic variance V_t and the asymptotic bias $B_I(t)$. Similarly to the formal analysis developed in subsection 2.3, we can capture this trade off by considering that volatility is so highly persistent that a block local to unity model may be relevant. In this case, as seen in subsection 2.2, we will have:

$$B_I(t) = O_P\left(1/\sqrt{n}\right) \tag{3.9}$$

so that $1/\omega_{n,t}^* = O_P(1)$.

So far we presented the limit theorems and main results in terms of the infeasible estimators. There are various ways this can be converted into a feasible limit theory. For example, in the absence of jumps a feasible asymptotic distribution is obtained by replacing V_t with a sample equivalent, namely, $\hat{V}_{n,t} = 2\sum_{t-1<t_{i+1}\leq t}(X_{n,t_{i+1}} - X_{n,t_i})^4$. Improved estimation of integrated quarticity can be derived in the spirit of section 2.4 (see Mykland and Zhang (2009)).[5]

Besides the estimation of integrated quarticity, we face another problem, namely the estimation of the forecast bias. We do not only observe the infeasible bias $B_I(t)$, but we even do not observe the feasible bias $B_{F,n}(t)$. The only thing we can actually compute is the following sample counterpart of $B_{F,n}(t)$:

$$\hat{B}_{F,n}(t) = \hat{\Theta}_{n,t} - \hat{\varphi}_{n,T}\hat{\Theta}_{n,t-1} - \hat{c}_{n,T} \tag{3.10}$$

[5]It should also parenthetically be noted that V_t can be estimated via several methods, including bootstrap proposed by Gonçalves and Meddahi (2009), subsampling by Kalnina and Linton (2007) and Kalnina (2011), or the observed AVAR approach of Mykland and Zhang (2017) as well as the examples discussed in section 7 of the latter paper.

where $\hat{\varphi}_{n,T}$ and $\hat{c}_{n,T}$ are OLS estimators in the time series regressions: $\hat{\Theta}_{n,t} = c_n + \varphi_n \hat{\Theta}_{n,t-1} + u_{n,t}$, $t = 1, \ldots, T$.

Note that the estimator (3.10) is essentially identical to the estimator considered in subsection 2.2, which implies that our implied estimated optimal weights $\hat{\omega}^*_{n,t}$ should be of sensible use, under similar assumptions regarding sample sizes n and T as well as volatility persistence (block local to unity). Strict application of the formulas developed in subsection 2.2 suggests the following natural sample counterpart of (3.7): $(\hat{\omega}^*_{n,t})^{-1} = 1 + [\hat{\varphi}^2_{n,T}\hat{V}_{n,t-1}]/\hat{V}_{n,t} + n\left[\hat{B}_{F,n}(t)\right]^2/\hat{V}_{n,t}$. However, it is worth keeping in mind that our block local to unity framework leads us to consider that:

$$1 - (\varphi^0)^2 = O(1/n).$$

In these circumstances, there is no point to use the estimator $\hat{\varphi}^2_{n,T}$ for the estimated optimal weights since the associated estimation error is of the same order of magnitude as the distance to unity. This is the reason why our preferred (approximated) optimal weight will be throughout:

$$\frac{1}{\hat{\omega}^*_{n,t}} = 2 + \frac{\hat{V}_{n,t-1} - \hat{V}_{n,t}}{\hat{V}_{n,t}} + n\frac{\left[\hat{B}_{F,n}(t)\right]^2}{\hat{V}_{n,t}}. \tag{3.11}$$

As noted before, the optimal weights are time varying. This sets our analysis apart from previous work only involving time invariant, or unconditional weighting schemes.[6] The comparison with unconditional schemes will be discussed at length in the next section. The fact that V_t is a stationary process implies that ω^*_t is stationary as well. It is also worth noting that the weight increases with V_t (relative to V_{t-1}). This is also expected as the measurement error is determined by V_t. High volatility leads to high V_t in fact. Hence, on high volatility days we expect to put more weight on the past to extract volatility.

To conclude, it should be noted that so far we confined our analysis to projections on one lag. We may consider higher order projections. In particular, it may be useful to think of the Beveridge-Nelson representation to accommodate the local-to-unity asymptotics (see e.g. Phillips and Solo (1992, Lemma 2.1) among others). We leave this for future research.

Finally, in Online Appendix Section OA.2 we discuss some additional volatility estimator

[6] In the context of volatility forecasting, Bollerslev, Patton, and Quaedvlieg (2016) have recently proposed time-varying weights that are pretty much conformable with our strategy of increasing the weight $\omega^*_{n,t}$ of past observation when current estimation error V_t is high with respect to the former one.

examples, including Two-Scales Realized Volatility and Multi-Scale and Kernel Realized Volatility, among others. Moreover, in Online Appendix Section OA.3 we discuss a set of examples which go beyond measures of quadratic variation, in particular: (a) estimation of Covariance from Asynchronous Observations, (b) block estimation of higher powers of volatility, and high frequency regression, and ANOVA.

3.2 The case with leverage effect, and nonzero μ

We consider the question of how to build a theory in the case where there is leverage effect. This means that the caveats at the beginning of Section 3.1 can be removed. In particular Assumptions 3.1 and 3.2 are replaced by:

Assumption 3.3 *We suppose that X_s is an Itô-semimartingale, either without jumps*

$$dX_s = \mu(s)ds + \sigma(s)dW(s),$$

or with jumps removed (see Footnote 2 in Section 3.1). The drift μ can vary quite freely (see Remark 1 below), and we allow for dependence between $\sigma(s)$ and (a) the jump sizes on the one hand, and (b) the driving Brownian motion and Poisson process on the other.

Assumption 3.4 *As $n \to \infty$:*

(i) The estimator $\hat{\Theta}_{n,t}$ is consistent;

(ii) A CLT applies, namely $n^{\alpha}(\hat{\Theta}_{n,t} - \Theta_t) \xrightarrow{\mathcal{L}} V_t^{1/2} N(0,1)$ stably in law, where V_t is a (potentially random) asymptotic variance, and where V_t is adapted to the filtration $\mathcal{F}_t$, which represents the history of underlying processes, including X, σ and θ.

In the no-leverage case, i.e. under Assumptions 3.1-3.2, one can condition on the σ_t process and then find the optimal estimator in terms of mean squared error. In the case with leverage, i.e. under Assumption 3.4, there is no general way of doing the conditioning for a fixed sample size. However, we show below that the asymptotic MSE (conditionally on the data, where the conditioning is done *after* the limit-taking) only depends on the σ_t process. The post-limit conditioning, therefore, gives rise to exactly the same formula that comes out

of the quite different procedure used in the no-leverage case. Therefore stable convergence saves the no-leverage result for the general setting.

The approach below can be used in many other settings, see, in particular, the results on estimation of the leverage effect in Example 4.

Remark 1 (Nonzero μ.) Once one is in the domain of stable convergence, all arguments go through for locally bounded drift process $\mu(s)$, cf. Mykland and Zhang (2009, p. 1407-1409). We can therefore assume that $\mu \equiv 0$ in Assumption 3.1. □

To proceed without conditioning, the point of departure is that the convergence in Assumption 3.2 remains valid even under leverage effect, as documented in our examples in Section 3.1. The underlying filtration is generated by a p-dimensional local martingale ($\chi^{(1)}$, ..., $\chi^{(p)}$). It is then the case that

$$n^{\alpha/2}(\hat{\Theta}_{n,t} - \Theta_t) \xrightarrow{d} Z_t\sqrt{V_t}, \tag{3.12}$$

where Z_t is standard normal, and the convergence is joint with ($\chi^{(1)}$, ..., $\chi^{(p)}$) (where this is a constant sequence). Z_t is independent of ($\chi^{(1)}$, ..., $\chi^{(p)}$) (the latter also occurs in the limit, since the sequence is constant as a function of n). This is known as stable convergence, see the papers cited, and also Rényi (1963), Aldous and Eagleson (1978), and Hall and Heyde (1980). It permits, for example, V_t to appear in the limit, while being a function of the data. As discussed in Section 5 of Zhang, Mykland, and Aït-Sahalia (2005), the convergence also holds jointly for days $t = 0, \ldots, T$. In this case, $Z_0, \ldots, Z_T$ are i.i.d.

With the convergence appearing in (3.12) in hand, one can now condition the asymptotic distribution on the data (*i.e.*, ($\chi^{(1)}$, ..., $\chi^{(p)}$)), and obtain that $Z_t\sqrt{V_t}$ is (conditionally) normal with mean zero and variance V_t. One can then develop the further theory based on asymptotic rather than small sample variances and covariances. Recall that it is convenient to make the persistence a function of n, hence $\varphi(n)$ (cf. equation (2.17) and the analogy with PMX). To distinguish small sample and asymptotic results, let us denote U_t by $U_t(n)$ as well and write:

$$\Theta_t = \varphi_n^0\Theta_{t-1} + \sqrt{1-(\varphi_n^0)^2}U_t(n) + (1-\varphi_n^0)E(\Theta_t). \tag{3.13}$$

One supposes that in the limit as $n \to \infty$, $\hat{\Theta}_{n,t-1}$ and $U_t(n)$ are uncorrelated. A similar,

feasible, equation is then written as

$$\hat{\Theta}_t = \varphi(n)\hat{\Theta}_{n,t-1} + \sqrt{1-\varphi(n)^2}\overline{U}_t(n) + (1-\varphi(n))E(\Theta_t). \tag{3.14}$$

Specifically, in analogy with (2.17),

$$n^\alpha(1-(\varphi_n^0)^2) = \gamma_0^2 \text{ and } n^\alpha(1-\varphi(n)^2) = \gamma^2. \tag{3.15}$$

Note that

$$1-\varphi(n) = n^{-\alpha}\frac{1}{2}\gamma^2(1+o_p(1)) = O_p(n^{-\alpha}). \tag{3.16}$$

Under the stationarity assumption, both $U_t(n)$ and $\overline{U}_t(n)$ have limits in law, which we denote by U_t and $\overline{U}_t$. If one subtracts (3.13) from (3.14), and then multiplies by $n^{\alpha/2}$, (3.12) yields

$$\begin{aligned} n^{\alpha/2}(\hat{\Theta}_{n,t}-\Theta_{n,t}) &= \varphi_n n^{\alpha/2}(\hat{\Theta}_{n,t-1}-\Theta_{n,t-1}) + n^{\alpha/2}(\varphi_n-\varphi_n^0)\Theta_{n,t-1} + n^{\alpha/2}\sqrt{1-\varphi_n^2}\overline{U}_{n,t} \\ &- n^{\alpha/2}\sqrt{1-(\varphi_n^0)^2}U_{n,t} - n^{\alpha/2}(\varphi_n-\varphi_n^0)E(\Theta_{n,t}) \to \sqrt{V_{t-1}}Z_{t-1} + \gamma\overline{U}_t - \gamma_0 U_t, \end{aligned} \tag{3.17}$$

whence

$$\gamma\overline{U}_t + \sqrt{V_{t-1}}Z_{t-1} = \gamma_0 U_t + \sqrt{V_t}Z_t, \tag{3.18}$$

hence U_t is not even asymptotically observable. Under this setup, from (3.16),

$$n^{\alpha/2}(\hat{\Theta}_{n,t} - \varphi(n)\Theta_{n,t-1}) \to \gamma\overline{U}_t. \tag{3.19}$$

Consider the best linear forecast of $\hat{\Theta}_{n,t}$ using (only) $\hat{\Theta}_{n,t-1}$: $\breve{\Theta}_{n,t|t-1} = \varphi_n\hat{\Theta}_{n,t-1} + (1-\varphi_n)E(\Theta)$, so that from (3.16)

$$n^{\alpha/2}(\hat{\Theta}_{n,t} - \breve{\Theta}_{n,t|t-1}) \to \gamma\overline{U}_t.$$

The final estimate is now $\hat{\Theta}_{n,t}(\omega_t) = \hat{\Theta}_{n,t} - \omega_t(\hat{\Theta}_{n,t} - \breve{\Theta}_{n,t|t-1})$, hence

$$\begin{aligned} n^{\alpha/2}(\hat{\Theta}_{n,t}(\omega_t) - \Theta_{n,t}) &= n^{\alpha/2}(\hat{\Theta}_{n,t} - \Theta_{n,t}) - \omega_t n^{\alpha/2}(\hat{\Theta}_{n,t} - \breve{\Theta}_{n,t|t-1}) \\ &\to \sqrt{V_t}Z_t - \omega_t\gamma\overline{U}_t = (1-\omega_t)\sqrt{V_t}Z_t - \omega_t\left[-\sqrt{V_{t-1}}Z_{t-1} + \gamma_0 U_t\right] \end{aligned} \tag{3.20}$$

by (3.18). Hence, the asymptotic mean squared error (AMSE; conditional on the data; the conditioning is done after taking the limit, which is the sequence consistent with stable

convergence) is $AMSE = (1-\omega_t)^2 V_t + \omega_t^2 \left[V_{t-1} + \gamma_0^2 U_t^2\right].$

The stable convergence (3.12) remains valid even in this triangular array setup (where $\Theta_{n,t}$ depends on n). For realized volatility, this is explicitly proved in Mykland and Zhang (2006, p. 1951-1952). For our other examples, the validity in the triangular array setting follows because all the cited papers use martingale central limit theorems which remain valid for triangular arrays.[7]

Using equation (3.13), one can therefore do the same calculations as before, but on asymptotic quantities. The asymptotic mean squared error (conditional on the data) of the overall estimate $QV_t(\omega_t)$ is minimized by $\omega_t^* = V_t/[V_t + V_{t-1} + \gamma_0^2 U_t^2]$. This is the same expression as equation (3.4). The further development is the same as in the no-leverage case.

3.3 Realized Betas

Another application of our theory would be to the case of realized betas, see in particular Barndorff-Nielsen and Shephard (2004a), Mykland and Zhang (2009, Section 4.2), and Aït-Sahalia, Kalnina, and Xiu (2020). To streamline the discussion we assume that there are equidistant synchronous observations, no microstructure noise, and no jumps (hence no need for truncation). Absent jumps, the methodologies of the two latter papers are similar; both use non-overlapping blocks, the former fixed size blocks, the latter asymptotically increasing size blocks. We here use a fixed block size M.[8] Mykland and Zhang (2009) provides a small sample interpretation under a contiguous probability distribution.

We are concerned with processes $X_{t,s}^{(1)}, ..., X_{t,s}^{(p)}$ and $Y_{t,s}$ observed on each day t at equidistant times $0 = t_{n,0} < t_{n,1} < ... < t_{n,n_1} = \mathcal{T}$, and related through

$$dY_{t,s} = \sum_{i=1}^{p} \beta_{t,s}^{(k)} dX_{t,s}^{(k)} + dZ_{t,s}, \text{ with } \langle X^{(k)}, Z\rangle_{t,s} = 0 \text{ for all } t,\ s \text{ and } k, \tag{3.21}$$

where we let s be time s on day t. To relate to the notation of the current paper, the

[7]See, for example, Theorem IX.7.19 or Theorem IX.7.28 of Jacod and Shiryaev (2003), Theorem B.4 of Zhang (2001) or Theorem 2.28 of Mykland and Zhang (2012). See also the books by Jacod and Protter (2012), and Aït-Sahalia and Jacod (2014). A summary discussion of technical tools can be found in Mykland and Zhang (2017, Section 7.2).

[8]The extension of the estimator to rolling blocks is straightforward, but would complicate the presentation given the already complex nature of the current setup.

integrated $\beta_{t,s}^{(k)}$ is $\Theta_t^{(k)} = \int_{t-1}^{t} \beta_{t,s}^{(k)}$. Blocks are defined by $\tau_{n,i-1} < t_{n,j} \leq \tau_{n,i}$,[9] where the $\tau_{n,i}$ correspond to every M^{th} observation time $t_{n,Mi}$. The estimator $(\hat{\beta}^{(1)}_{t,n,\tau_{n,i-1}}, ..., \hat{\beta}^{(p)}_{t,n,\tau_{n,i-1}})$ of $(\beta^{(1)}_{t,\tau_{n,i-1}}, ..., \beta^{(p)}_{t,\tau_{n,i-1}})$ is the the regular least squares estimator (without intercept) based on the observables $(\Delta X^{(1)}_{t,n,t_{n,j}}, ..., \Delta X^{(p)}_{t,n,t_{n,j}}, \Delta Y_{t,n,t_{n,j}})$ inside the block from $\tau_{n,i-1}$ to $\tau_{n,i}$. The overall estimate of the vector of $\Theta^{(k)}_{n,t}$'s is then

$$\hat{\Theta}^{(k)}_{n,t} = \sum_i \hat{\beta}^{(k)}_{t,n,\tau_{n,i-1}} M\Delta t. \tag{3.22}$$

The asymptotic covariance matrix is given by $V_t = \frac{M\mathcal{T}}{M-p-1} \int_0^{\mathcal{T}} \langle Z, Z\rangle'_s (\langle X, X\rangle'_s)^{-1} dt$, for s-values on day t (so $Z = Z_{t,\cdot}$, and similarly for X), and where $\mathcal{T}$ is the notional length of the time period from $t-1$ to t, often one day. This is from Mykland and Zhang (2009, eq. (72), p. 1426). It is consistent with Aït-Sahalia, Kalnina, and Xiu (2020, Theorem 2, p. 91), with correction for fixed M and for normalization by $n^{1/2}$ rather than $\Delta t^{-1/2}$.

Depending on taste (and with the same outcome), the estimator of covariance is obtained from regression considerations, or as approximations to the asymptotic values. Let $RSS_{t,n,i}$ be the residual sum of squares in the (no-intercept) regression in block i on day t with sample size n, *i.e.*,

$$RSS_{t,n,i} \;=\; \sum_{\tau_{n,i-1} < t_{n,j} \leq \tau_{n,i}} (\widehat{\Delta Z}_{t,n,t_{n,j}})^2.$$

Similarly, let RV_i be the realized volatility (sum of squares) matrix in block # i. Since $RSS_{t,n,i} \approx \widehat{\langle Z, Z\rangle}_{t,\tau_{n,i}} - \widehat{\langle Z, Z\rangle}_{t,\tau_{n,i-1}}$ and $RV_{t,n,i} \approx \widehat{\langle X, X\rangle}_{t,\tau_{n,i}} - \widehat{\langle X, X\rangle}_{t,\tau_{n,i-1}}$ (or following the cited papers),

$$\widehat{V}_{t,n} \;=\; \frac{M\mathcal{T}}{M-p-1} M\Delta t \sum_i RSS_{t,n,i} {RV_{t,n,i}}^{-1}.$$

Once again, recall that there is a different block i for each time period t. The remaining coefficients follow from equations (3.10)-(3.11), where we recall that $\hat{\varphi}_{n,T}$ and $\hat{c}_{n,T}$ are OLS estimators in the time series regressions: $\hat{\Theta}_{n,t} = c_n + \varphi_n \hat{\Theta}_{n,t-1} + u_{n,t}$, $t = 1, \ldots, T$.

[9] On each day t. We have not marked the dependence of $\tau_{n,i-1}$ and $t_{n,j}$ on day t, to reflect an idealized situation where the times are the same on each day. In a more realistic situation, these quantities depend also on day number; we have avoided this here in the interest of simplicity of notation. On the other hand, $\hat{\beta}^{(1)}_{t,n,\tau_{n,i-1}}$ will differ depending on day t and sample size n, and is marked accordingly.

3.4 Ill-Posed Estimation Problems

We have so far considered the relatively well-posed problem of estimating volatility from high frequency data. The use of multi-day data, however, really comes even more into play when trying to estimate less well-posed quantities. In particular, we consider:

$$\begin{aligned} \text{the volatility of volatility:} \quad & \Theta_t = \langle \sigma, \sigma \rangle_t - \langle \sigma, \sigma \rangle_{t-1} \\ \text{the (instantaneous) leverage effect:} \quad & \Theta_t = \langle p, \sigma \rangle_t - \langle p, \sigma \rangle_{t-1}. \end{aligned} \tag{3.23}$$

Estimators of the volatility of volatility have been proposed by Mykland, Shephard, and Sheppard (2012), whereas instantaneous leverage effect estimators are covered by Wang and Mykland (2014). Such estimators can account for leverage effects at multiple horizons, and are closely related to news impact curves (Engle and Ng (1993), Chen and Ghysels (2011)). Nonparametric estimation of the leverage effect is developed in Wang and Mykland (2014).[10]

For brevity, we give a summary account of how our theory extends to such situations. The generalization is straightforward but the details would be notationally heavy. For simplicity, we consider the no-jump case, which can be undone as in Aït-Sahalia, Fan, Laeven, Wang, and Yang (2017). Specifically, $dW_t = \rho_t dW_{1t} + \sqrt{1-\rho_t^2} dW_{2t}$, and the system is given by dp_t $=\mu_t dt + \sigma_t dW_t$ and $d\sigma_t^2 = \nu_t dt + \gamma_t dW_{1t}$.

To define estimators for the parameters in (3.23), we need a block size M_n, determined by the econometrician, held to satisfy $n^{-1/2} M_n \to c$, as $n \to \infty$. Let $\tau_i = \tau_{n,t,i}$ be the iM_nth observation time on day t. The number of blocks on the form $(\tau_{i-1}, \tau_i]$ is given by $n_M = \lfloor n/M_n \rfloor$. For each block, define a local realized volatility as $RV_i = \sum_{\tau_{i-1} < t_j \leq \tau_i} (\Delta p_{t_j})^2$, $i = 1, \ldots, n_M$. In the absence of microstructure noise, $\hat{\sigma}^2_{\tau_{n,i}} = [M_n \times \Delta t]^{-1} RV_i$ is a consistent estimator of spot volatility (Foster and Nelson (1996), Comte and Renault (1998), Mykland and Zhang (2008)).

Example 3 Estimation of Volatility of Volatility (No Microstructure Noise) We follow Mykland, Shephard, and Sheppard (2012) (MSS), in particular their Section 6.[11]

[10] The concept of leverage effect is used to cover several concepts, cf the review in the Introduction to the cited papers.

[11] An alternative estimator can be found in (Vetter 2015).

Define *blocked bipower variation* and *edge corrected realized volatility* as, respectively,

$$BV_{M,n} = \frac{n}{(n_M - 1)\,(E(\chi^2_{M_n})^{1/2})^2} \sum_{i=2}^{n_M} (RV_{i-1} RV_i)^{1/2}$$

$$ECRV_n^{(2)} = \frac{1}{2} RV_1 \; + \; \sum_{i=2}^{n_M - 1} RV_i + \frac{1}{2} RV_{n_M}$$

where χ^2_M is a χ^2 random variable with M degrees of freedom. Define the intra-day based estimator as $\hat{\Theta}_{n,t} = \widehat{\langle \sigma, \sigma \rangle}^{(n)} = c^{-1} \frac{3}{K-1} n^{1/2} \left(ECRV^{(2)_n} - BV_{M,n} \right)$. Following Corollary 2 (Section 6.2) in MSS, (3.12) is now valid with $\alpha = 1/4$ and $V_t = \frac{27}{16} \int_{t-1}^{t} c^{-3} \left(2\sigma_s^2 + c^2 \langle \sigma, \sigma \rangle'_s \right)^2 ds$. The discussion in MSS also extends to estimators of $\langle \sigma, \sigma \rangle$ based on multipower variation. □

Example 4 Estimation of Leverage Effect (No Microstructure Noise) We follow Wang and Mykland (2014) (WM), in particular their Sections 2.2-2.3.[12] Define the intra-day based estimator as $\bar{\theta}_t = \widehat{\langle X, \sigma \rangle}^{(n)} = 2 \sum_{i=0}^{M_n - 2} (X_{\tau_{n,i+1}} - X_{\tau_{n,i}})((\hat{\sigma}^2_{\tau_{n,i+1}})^{1/2} - (\hat{\sigma}^2_{\tau_{n,i}})^{1/2})$. Following Theorem 1 (Section 2.3) in WM, (3.12) is now valid with $\alpha = 1/4$ and $V_t = \frac{4}{c} \int_{t-1}^{t} \sigma_t^4 dt + c \int_{t-1}^{t} \sigma_t^2 (\frac{44}{3} f_t^2 + \frac{32}{3} g_t^2) dt$ where $f_t = \gamma_t \rho_t$ and $g_t = \gamma_t (1 - \rho_t^2)^{1/2}$. □

Example 5 Estimation In the Presence of Microstructure. We refer to the more complex discussions in Sections 6.1.3-6.1.4 of MSS (Volatility of Volatility) and Section 4.1 of WM (Leverage Effect). With suitable definitions of Θ_t and V_t, (3.12) remains valid with $\alpha = 1/8$. □

As seen from our examples, the following is typically the case. When there is no microstructure, variances, covariances, regression coefficients, etc, can be estimated as rate $\alpha = 1/2$, whereas more ill-posed quantities (such as leverage effect and volatility of volatility) will at best be estimable at rate $\alpha = 1/4$. When there is microstructure noise, both of these rates are halved, to $\alpha = 1/4$ and $\alpha = 1/8$, respectively.

Recall that α is the convergence rate in Conditions [ii] in Assumption 3.2 and [ii] in Assumption 3.4. The closer α is to zero, the more ill-posed the problem, and hence the greater the

[12] WM's F function is here chosen as $F(x) = x^{1/2}$. Other choices of F provide estimators of skewness ($F(x) = x$) and regression of σ^2 on X ($F(x) = \log(x)$). In both these cases, our theory of across-day estimation remains valid.

gains of our proposed procedure. To see this replace Assumption (2.18) by

$$n^{2\beta}(1-(\varphi_n^0)^2)=\gamma_0^2 \text{ and } n^{2\beta}(1-\varphi(n)^2)=\gamma^2, \tag{3.24}$$

where we can think of β as measuring the persistence of the underlying system. Note that so far we have identified the two rates to get a balanced asymptotic expression.

As documented in our simulations, there is gain for volatility estimation from across-day inference, and so the choice $\beta = 1/2$ is meaningful for the most well-posed case when $\alpha = 1/2$. For the estimators with slower convergence, $\alpha < 1/2$, and with the same choice $\beta = 1/2$, the across-day information dominates the intraday information. Therefore, for ill-posed estimators, our theory is at its most effective. This is consistent with the comment in Wang and Mykland (2014, p. 205-207), to the effect that a day was thought to be a slightly too short time period to estimate leverage effect.

4 Alternative Weighting Schemes

The topic of this paper was originally considered in earlier work by Andreou and Ghysels (2002) who tried to exploit the continuous record asymptotic analysis of Foster and Nelson (1996) for the purpose of improving realized volatility measures. At the time the paper by Andreou and Ghysels (2002) was written the in-sample asymptotics was not taken into account by the authors, as their paper was concurrent to that of Barndorff-Nielsen and Shephard (while the early work of Jacod was discovered only much later). Therefore Andreou and Ghysels (2002) failed to recognize that increased accuracy of in-sampling will diminish the need to use past data. This does not occur in the context of Foster and Nelson (1996) who study instantaneous or spot volatility. In the latter case persistence will remain relevant to filter current spot volatility, which is the key difference between continuous record and in-sample asymptotics. An early draft of Meddahi (2002) included a section which discussed the same question and where it was recognized that optimal filter weights should depend on the in-sample frequency and ultimately become zero asymptotically. There are many important differences between the analysis in the current paper and the filtering approach pursued by Andreou and Ghysels and Meddahi. The most important difference is that we derive *conditional* filtering schemes, dependent on the path of the volatility process, whereas Andreou and Ghysels and Meddahi only consider unconditional, that is time-invariant, filtering.

In this section we walk step-by-step from the unconditional to the optimal model-free weighting scheme we introduced in the previous section. We start by noting that equations (3.10)-(3.11) suggest two feasible weighting schemes: $(\omega_{vt}^*)^{-1} = 2 + .5[n^{2\alpha}Var(\hat{\Theta}_t - \hat{\Theta}_{n,t|t-1})]/\hat{V}_{n,t} + [\hat{V}_{n,t-1} - \hat{V}_{n,t}]/\hat{V}_{n,t} = 2 + .5[n^{2\alpha}Var(U_t)]/\hat{V}_{n,t} + [\hat{V}_{n,t-1} - \hat{V}_{n,t}]/\hat{V}_{n,t}$, where we use the v subscript to refer to the unconditional variance of U, and the subscript u relates to the actual value of U_t, yielding: $(\omega_{ut}^*)^{-1} = 2 + .5[n^{2\alpha}(\hat{\Theta}_t - \hat{\Theta}_{n,t|t-1})^2]/\hat{V}_{n,t} + [\hat{V}_{n,t-1} - \hat{V}_{n,t}]/\hat{V}_{n,t} = 2 + .5[n^{2\alpha}(\overline{U}_t)^2]/\hat{V}_{n,t} + [\hat{V}_{n,t-1} - \hat{V}_{n,t}]/\hat{V}_{n,t}$. When $\alpha = 1/2$, in the case of quadratic variation, we have:

$$(\omega_{vt}^*)^{-1} = 2 + \frac{nVar(\hat{\Theta}_{n,t} - \hat{\Theta}_{n,t|t-1})}{2\hat{V}_{n,t}} + \frac{\hat{V}_{n,t-1} - \hat{V}_{n,t}}{\hat{V}_{n,t}} \tag{4.1}$$

$$(\omega_{ut}^*)^{-1} = 2 + \frac{n(\hat{\Theta}_{n,t} - \hat{\Theta}_{n,t|t-1})^2}{2\hat{V}_{n,t}} + \frac{\hat{V}_{n,t-1} - \hat{V}_{n,t}}{\hat{V}_{n,t}} \tag{4.2}$$

where $\hat{V}_{n,t} = n(\pi/2)^2 \sum_{j=4}^{n} |X_{n,t_j} X_{n,t_{j-1}} X_{n,t_{j-2}} X_{n,t_{j-3}}|$ in the case of BPV. The distinction between (4.1) and (4.2) will be important when we turn to processes with leverage effect, as will be discussed in Online Appendix Section 3.2.

These two weighting schemes will be compared with unconditional schemes at first and a conditional scheme, while sub-optimal, provides a natural link between the unconditional and conditional schemes (4.2) and (4.1). To discuss unconditional weighting schemes we drop time subscripts to the weights ω_t^* in equation (3.4) and consider the generic class of estimators:

$$\hat{\Theta}_{n,t}(\omega) = \hat{\Theta}_{n,t} - \omega(\hat{\Theta}_{n,t} - \hat{\Theta}_{n,t|t-1}). \tag{4.3}$$

Recall that $\int_{t-1}^{t} \theta(s)ds$ - $\Theta_{t|t-1} = \sqrt{1-(\varphi_n^0)^2}U_t$ and that the relevant trade-off is captured by the product $n^\alpha(1-(\varphi_n^0)^2)$, which resulted in the local-to-unity asymptotics. The analysis of Andreou and Ghysels (2002) did not recognize these trade-offs and it is perhaps useful to start with their rule-of-thumb approach which consisted of setting $\varphi_0 = \varphi = 1$, de facto a unit root case, and therefore $\breve{\Theta}_{n,t|t-1} = \hat{\Theta}_{n,t-1}$. Moreover, their approach is unconditional and therefore overlooks the difference between V_t and V_{t-1}. The unit root case yields the weighting scheme $\omega^{r-th} = .5$ (substituting $\varphi_0 = \varphi = 1$ in equation (4.2) with $V_t = V_{t-1}$), and the rule-of-thumb estimator:

$$\hat{\Theta}_{n,t}(\omega^{r-th}) = .5\hat{\Theta}_{n,t} + .5\hat{\Theta}_{n,t-1}. \tag{4.4}$$

This is the first of two unconditional weighting schemes. Unlike Andreou and Ghysels (2002), Meddahi recognized the trade-off captured by the product $n^\alpha(1-(\varphi_n^0)^2)$, and constructed a *model-based* weighting scheme, denoted by ω^{unc} and which is characterized as $\omega^{\text{unc}} = [2+2\lambda]^{-1}$ with:

$$\lambda = n^\alpha[1-\varphi_n^2]\frac{Var[\Theta_{t|t-2}]}{E(V_t)} \simeq \frac{\gamma^2}{4}\frac{Var[\Theta_{t|t-1}]}{E(V_t)}. \tag{4.5}$$

It should be noted that Meddahi used the unconditional variance of the estimation error of quadratic variation, that is using our notation $E(V_t/2)/n^\alpha$.[13] Moreover, he assumed an explicit data generating process to compute the weights, hence a model is needed to be specified (and estimated) to compute the weights.[14] To obtain a feasible scheme, we will use unconditional *sample* means of Θ_t and V_t. In this respect we deviate from the model-based approach of Meddahi, namely we do not use any explicit model to estimate the weighting schemes. The above derivation suggests the second unconditional scheme:

$$(\omega^{unc})^{-1} = 2 + \frac{nVar(\Theta_t - \Theta_{t|t-1})}{2E(V_t)} = 2 + \gamma^2\frac{Var[U_t]}{2E(V_t)},$$

which again does not depend on t, and where in practice the term $(\gamma^2 Var[QV])/E(Q)$ will be computed as the ratio of $(1-\varphi^2)Var[QV]$ and the asymptotic (unconditional) variance $E(Q/(n^\alpha))$ of the estimation error for integrated volatility.

To appraise the differences between conditional and unconditional weighting schemes we also compare:

$$(\omega^{unc})^{-1} \simeq 2 + \gamma^2\frac{Var[U_t]}{2E(V_t)} \tag{4.6}$$

with our optimal estimator slightly rewritten as:

$$[\omega_t^*]^{-1} \simeq (2 + \gamma^2\frac{Var(U_t)}{2V_t}) + HV_t \tag{4.7}$$

where $HV_t = (\hat{V}_{n,t-1} - \hat{V}_{n,t})/\hat{V}_{n,t}$ which we referred to as a heteroskedasticity correction. An approach intermediate between (4.7) and (4.6), that is unconditional up to heteroskedasticity

[13] It should be noted that Andersen, Bollerslev, and Meddahi (2005, footnote 9) provided a solution for a model-free estimation.

[14] To clarify the difference between our model-free approach and Meddahi, it should be noted that the weights in our analysis are *not* based on a specific model. Moreover, the prediction formula in our analysis is nothing but a projection and does not involve any AR model.

Table 1: Inverse weighting schemes

	QV	BPV
$(\omega^{unc})^{-1}$	$2+\gamma^2\frac{Var[U_t]}{2E(V_t)}$	$2+\frac{nVar[\hat{\Theta}_{n,t}-\hat{\Theta}_{n,t\|t-1}]}{\nu E(V_t)}$
$(\omega_t^{hc})^{-1}$	$2+HV_t+\gamma^2\frac{Var[U_t]}{2E(V_t)}$	$2+\frac{nVar[\hat{\Theta}_{n,t}-\hat{\Theta}_{n,t\|t-1}]}{\nu E(V_t)}+\frac{\hat{V}_{n,t-1}-\hat{V}_{n,t}}{\hat{V}_{n,t}}$
$(\omega_{vt}^{*})^{-1}$	$2+HV_t+\gamma^2\frac{Var[U_t]}{2V_t}$	$2+\frac{nVar[\hat{\Theta}_{n,t}-\hat{\Theta}_{n,t\|t-1}]}{\nu\hat{V}_{n,t}}+\frac{\hat{V}_{n,t-1}-\hat{V}_{n,t}}{\hat{V}_{n,t}}$
$(\omega_{ut}^{*})^{-1}$	$2+HV_t+\gamma^2\frac{\overline{U}_t^2}{2V_t}$	$2+\frac{n(\hat{\Theta}_{n,t}-\hat{\Theta}_{n,t\|t-1})^2}{\nu\hat{V}_{n,t}}+\frac{\hat{V}_{n,t-1}-\hat{V}_{n,t}}{\hat{V}_{n,t}}$

Notes: In addition to the above weighting schemes $\omega^{r-th}=1/2$. Moreover, in the above $\overline{U}_t=\hat{\Theta}_{n,t}$ - $\hat{\Theta}_{n,t-1}$.

correction yields:

$$(\omega_t^{hc})^{-1}=2+\frac{\hat{V}_{n,t-1}-\hat{V}_{n,t}}{\hat{V}_{n,t}}+\gamma^2\frac{Var[U_t]}{E(V_t)}=2+HV_t+\gamma^2\frac{Var[U_t]}{2E(V_t)}. \tag{4.8}$$

The above weighting schemes provide a progression towards the optimal (conditional) weighting scheme. Starting with the rule-of-thumb scheme ω^{r-th}, we progress to ω^{unc} where unconditional moments are used, followed by the heteroskedasticity correction embedded in ω_t^{hc}. To summarize, we have five possible weights, focusing only on QV and BPV appearing in Table 1.

From the above analysis we can make several observations:

- The unconditional formula of Meddahi gives a weight to past realized volatility smaller than the rule-of-thumb weight of (1/2).[15]
- This unconditional formula does not take into account the conditional heteroskedasticity that is due to the (asymptotic) estimation error of realized volatility. For instance, when $\hat{V}_{n,t}>\hat{V}_{n,t-1}$ that is a larger estimation error on current integrated volatility estimation than on the past one, we may be led to choose a weight larger than (1/2) for past realized volatility. Typically, taking the term HV_t into account should do better in the same way WLS are more accurate than OLS in case of conditional heteroskedasticity.

[15]Moreover, as noted before, the weight diminishes with n.

- Besides the heteroskedasticity correction HV_t we also observe that $Var[U_t]/E(V_t)$ is replaced by $Var[U_t]/V_t$ in optimal weighting schemes.

Finally, in the same spirit as the rule-of-thumb approach we also consider a common industry practice, known as *RiskMetrics,* which is characterized as:

$$\hat{\Theta}_{n,t}(\omega^{rm}) = \omega^{rm}\hat{\Theta}_{n,t} + (1-\omega^{rm})\hat{\Theta}_{n,t-1} \tag{4.9}$$

where $|\omega^{rm}| < 1$. Note that the Andreou and Ghysels rule-of-thumb estimator is a special case with $\omega^{rm} = .5$. We will consider $\omega^{rm} = .9$ in our analysis, although we also experimented with slightly higher and low values in the $[.8, .95]$ range.

5 A simulation study

The purpose of the simulation is two-fold. First, we want to assess the efficiency gains of the optimal weighting schemes. This will allow us to appraise how much can be gained from filtering. Second, we would like to compare the feasible optimal weighting schemes ω_t^* with the rule-of-thumb scheme ω^{r-th}, the unconditional scheme ω^{unc} and the heteroskedasticity correction embedded in ω_t^{hc}. This will allow us to appraise the difference between conditional and unconditional filtering as well as the relative contribution of the natural progression towards the optimal (conditional) starting with the rule-of-thumb scheme ω^{r-th}, to ω^{unc}, followed by the heteroskedasticity correction embedded in ω_t^{hc}. Finally, we also consider the *RiskMetrics* approach featuring ω^{rm} discussed in equation (4.9). While the simulations used empirically plausible data generating processes, we also conducted a small empirical study that yielded similar results. An extensive empirical study is beyond the scope of the present paper and left for future research.

5.1 Simulation design

We consider 1,000 replications of samples each consisting of 500 and 1,000 'days' with in-sample (intra-daily sampling) sizes $n = 288$ and 144 (following a first 500 days burn-in pre-sample). These correspond to the use of five-minute and ten-minute returns in a 24-hour financial market. We treat the one-minute quantities as the "truth", hence they provide us

with a benchmark for comparison. Every simulation has a 1000 days burn-in pre-sample period to eliminate starting value problems.

The simulations pertain to respectively (a) volatility and (b) betas. A subsection is devoted to each separately.

5.1.1 Volatility

The first class of models we simulate is based on Andersen, Bollerslev, and Meddahi (2005) and consists of:

$$d\log S_t = \mu dt + \sigma_t \left[\rho_1 dW_{1t} - \rho_2 dW_{2t} + \sqrt{1-\rho_1^2-\rho_2^2} dW_{3t}\right]. \tag{5.1}$$

When $\mu = \rho_1 = \rho_2 = 0$, we obtain:

$$d\log S_t = \sigma_t dW_{3t}. \tag{5.2}$$

The dynamics for the instantaneous volatility is one of the following (with the specific parameter values taken from Andersen, Bollerslev, and Meddahi (2005)):

$$d\sigma_t^2 = .035(.636 - \sigma_t^2)dt + .144\sigma_t^2 dW_{1t}, \tag{5.3}$$

which is a GARCH(1,1) diffusion, or a two-factor affine model:

$$\begin{aligned} d\sigma_{1t}^2 &= .5708(.3257 - \sigma_{1t}^2)dt + .2286\sigma_{1t}^2 dW_{1t}, \\ d\sigma_{2t}^2 &= .0757(.1786 - \sigma_{2t}^2)dt + .1096\sigma_{2t}^2 dW_{2t}, \end{aligned} \tag{5.4}$$

with $\sigma_t^2 = \sigma_{1t}^2 + \sigma_{2t}^2$. All of the above models satisfy the regularity conditions of the Jacod (1994) and Barndorff-Nielsen and Shephard (2002b) asymptotics. We also considered cases with nonzero μ, ρ_1, and/or ρ_2. Hence, these are diffusions with drift and leverage. We do not report those results as they were similar to the no-drift/leverage results.

We also include a third class of processes involving jump diffusions based on Huang and

Tauchen (2006). They analyze a stochastic volatility jump diffusion model, labeled as SV1FJ:

$$\begin{aligned} d\log S_t &= \mu dt + \exp(\beta_0 + \beta_1\sigma_t)(\rho dW_{1t} + (1-\rho)dW_{2t} + dL_t, \\ d\sigma_t &= \alpha_v dt + dW_{2t}, \end{aligned} \tag{5.5}$$

where L_t is a Compound Poisson Process with constant jump intensity λ and random jump size distributed as $N(0, \sigma^2_{jmp})$. The model SV1FJ has one stochastic volatility factor and a jump, hence the acronym. The parameter values are taken from Huang and Tauchen (2006, Table 1) and are based on prior empirical results, most notably Chernov, Gallant, Ghysels, and Tauchen (2002) and Andersen, Benzoni, and Lund (2002). In particular, we select the empirically most plausible values from Huang and Tauchen (2006, Table 1), and set $\alpha_v = -0.1$ and $\lambda = 0.058$ and $\sigma^2_{jmp} = 1.50$. The simulation of the jump process is exactly as it is described in Huang and Tauchen (2006).[16]

For the continuous-path process simulations we consider weights for QV, with $(\omega^*_{vt})^{-1}$ appearing in equation (4.1), $(\omega^*_{ut})^{-1}$ appearing in equation (4.2), and $(\omega^{hc}_t)^{-1}$ from equation (4.8), and $(\omega^{unc})^{-1}$ from equation (4.6). For the jump diffusion we consider the BPV weighting schemes as specified in Table 1.

Finally, for the purpose of constructing the weights, AR(1) prediction schemes are used. Hence, all our simulations consider looking at the previous day's QV only. To assess estimation uncertainty we also consider cases where $(\omega^*_{vt})^{-1}$, $(\omega^{hc}_t)^{-1}$ and $(\omega^{unc})^{-1}$ are based on sample moments which are perturbed by extra noise.

5.1.2 Betas

Our final simulation design pertains to CAPM betas. We start from equation (3.21) and assume that the single factor is the market return process which we assume to obey the law of motion of the process appearing in equation (5.2) with instantaneous volatility dynamics appearing in (5.4).

To complete the model we need to specify the beta dynamics. Here we follow Aït-Sahalia, Kalnina, and Xiu (2020) and use an Ornstein-Uhlenbeck process for the time variation of the market betas, namely:

$$d\beta_t = \kappa(\mu - \beta_t)dt + \sigma_b dW_t, \tag{5.6}$$

[16] We are grateful to Xin Huang for sharing the simulation code with us.

where we take the parameter values similar to those in Aït-Sahalia, Kalnina, and Xiu (2020, Table 1). In particular, we take for DGP1 $\kappa = 2$, $\mu = 1.0$, with $\sigma_b = 3$, for DGP2 $\kappa = 2$, $\mu = 0.10$ and $\sigma_b = 0.03$, for DGP3 $\kappa = 1$, $\mu = 1.0$ and $\sigma_b = 0.03$. The DGPs are different with respect to (a) persistence, i.e. κ, (b) unconditional betas, i.e. μ and (c) noise of the innovations represented by σ_b. The CAPM equation has no intercept, meaning it pertains to excess returns, and the idiosyncratic error variance equals one. All returns are sampled at 1 and 5 minute intervals with sample size $T = 500$.[17] The weighting schemes are those appearing in Table 2.

5.2 Simulation results

Table 2 reports the first set of simulation results. The table consists of three panels, corresponding to respectively the GARCH diffusion, two-factor SV diffusion and Jump diffusion. For each we report the mean of the gains in MSE relative to a standard QV or BPV estimator using the same frequency data. The results in the table pertain to 5 and 10 minute sampling. We focus our analysis in particular on the 10 minute sampling as this is arguably the most relevant one regarding across-sample efficiency gains. We consider three time-varying weighting schemes, namely those driven by ω^*_{ut}, ω^*_{vt} and ω^{hc}_t, and the fixed weighting schemes, namely ω^{unc} and ω^{rm}. We also report the mean and variance of the weights obtained from the simulations. We report the results for $T = 500$ as they are representative for the two sample sizes ($T = 500$ and 1000) we considered.

We observe that overall the weighting schemes based on ω^*_{ut} and ω^*_{vt} appear to be consistently among the best weighting schemes in terms of MSE improvement across all three DGPs considered and the difference between schemes ω^*_{ut} and ω^*_{vt} is insignificant. It is worth recalling that ω^*_{vt} is more closely related to what would be the Kalman filter weights.

The results for the GARCH diffusion indicate that the optimal weighting scheme is the best, in terms of average MSE improvements. The two-factor diffusion and jump diffusion cases yield less impressive MSE gains for the optimal weighting schemes, and the latter are particularly challenged by the *RiskMetrics* approach which in both cases does either marginally better (two-factor DGP) or substantially better (jump-diffusion). The other weighting schemes do not perform very well in comparison to the optimal and the *RiskMetrics* one.

[17] We also considered $T = 1000$ but are not reported due to the similarity to those reported here.

Finally, we turn to Table 3 where we report the results pertaining to realized betas. The entries are the ratios of MSE of the various estimators vis-à-vis the MSE of a standard realized beta and the mean of the weights. The gains in terms of MSE improvements are much more substantial this time compared to the volatility examples - particularly the two-factor SV and jump diffusion cases. Starting with the 5 minute sampling we see gains of about 25 % for the first DGP and 20 % when we look at the high volatility of beta innovations in the third GDP. Not surprisingly, we note from the second DGP that the level of beta has very little impact. Among the weighting schemes we note that ω_{vt}^* appears to be consistently the best, with only small differences between ω_{ut}^*, ω_t^{hc}, and ω^{unc}. It is interesting to note that even at the 1 minute sampling scheme we still see gains that are substantial. The mean of the weights are as high as 20%, as reported in Panel B of the table, for the 5 minute sampling scheme.

6 Conclusions

We revisited the widely used in-sample asymptotic analysis extensively used in the realized volatility literature and showed that there are gains to be made in estimating current realized volatility from considering realizations in prior periods. The main focus on the paper was establishing the theory and showing its potential importance. There are still many implications of our results for hypothesis testing which were not covered in our paper. In particular, discriminating between jumps-diffusions and diffusions, i.e. testing for the presence of jumps is an important example. Since such tests rely on various data-driven high frequency statistics, including higher order moment-based ones, there is scope for improving their sampling properties with our approach. We leave this topic for future research.

Appendix

A A Comparison of Two Estimators

The unbiased estimator defined in equation (2.20) can be rewritten as,

$$\hat{\sigma}^{[4]}_{n,t} = \frac{m}{n}\sum_{t=1}^{m/n}\frac{\hat{\sigma}^4_{n,t,[i]}}{1+2/m} = \frac{n}{m+2}\sum_{i=1}^{n/m}[\sum_{j=m(i-1)+1}^{j=mi} X^2_{t,j}]^2. \tag{A.1}$$

The above estimator can be compared with the naive estimator appearing in (2.21). To do so we need to derive the conditional variance of $\hat{\sigma}^{[4]}_{n,t}$. Note that we can rewrite the estimator (2.20) as:

$$\hat{\sigma}^{[4]}_{n,t} = \frac{1}{(m+2)n}\sum_i \sigma^4_{n,t,[i]}[\sum_j(\frac{X_{t,j}}{\sigma_{n,t,[i]}/\sqrt{n}})^2]^2 = \frac{1}{(m+2)n}\sum_i \sigma^4_{n,t,[i]}[\chi^2_i(m)]^2.$$

Since $E\left[[\chi^2_i(m)]^{p/2}\right] = 2^{p/2}\Gamma((p+m)/2)/\Gamma(m/2)$, therefore $E[\chi^2_i(m)]^4 = 2^4\ \Gamma(4+m/2)/\Gamma(m/2) = 2^4\ (3+m/2)\ (2+m/2)\ (1+m/2)\ m/2$. Consequently, $E[\chi^2_i(m)]^4 = (m+6)(m+4)(m+2)m$. Along similar lines, one has $E[\chi^2_i(m)]^2 = m(m+2)$. Therefore, $Var[\chi^2_i(m)^2] = (m+6)(m+4)(m+2)m$ - $m^2(m+2)^2 = 8m(m+2)(m+3)$. This yields:

$$Var_\sigma[\hat{\sigma}^{[4]}_{n,t}] = \frac{1}{n^2(m+2)^2}\sum_i \sigma^8_{n,t,[i]}8m(m+2)(m+3) = \frac{8m(m+3)}{n^2(m+2)}\sum_i \sigma^8_{n,t,[i]}.$$

We now turn our attention to the naive estimator written as in equation (2.21):

$$\tilde{\sigma}^{[4]}_{n,t} = \frac{1}{3n}\sum_i \sigma^4_{n,t,[i]}\sum_j(\frac{X_{t,j}}{\sigma_{n,t,[i]}})^4 = \frac{1}{3n}\sum_i \sigma^4_{n,t,[i]}\sum_j(\chi^2(1))^2.$$

Therefore, $Var(\tilde{\sigma}^{[4]}_{n,t}) = \frac{1}{9n^2}\sum_i \sigma^8_{n,t,[i]}Var((\chi^2(1))^2)\times m = \frac{32m}{3n^2}\sum_i \sigma^8_{n,t,[i]}$. From these results we can deduce that there will be efficiency improvements provided that $(m+3)/(m+2) < 4/3$, or $3m+9 < 4m+8$, that is $m > 1$.

To conclude, we compute the unconditional variance of $\hat{\sigma}^4_{n,t}$. First, note that

$$\hat{\sigma}^4_{n,t} = \frac{1}{n(m+2)}\sum_{i=1}^{m/n}\sigma^4_{n,t,[i]}\varepsilon^2_i \tag{A.2}$$

with $\varepsilon_i^2 \sim \chi^2(m)$. Therefore the unconditional variance of $\hat{\sigma}_{n,t}^4$ can be written as:

$$Var[\hat{\sigma}_{n,t}^4] = \frac{n^2}{n^2(n+2/n)^2} \sum_{i=1}^{m/n} Var[\sigma_{n,t,[i]}^4](E[\varepsilon_i^2])^2 + 2(Var[\varepsilon_i^2])E[\sigma_{n,t,[i]}^8]. \tag{A.3}$$

Given the definition of ε_i^2, we have that $E\varepsilon_i^2 = 1 + 2/\mathrm{m}$, and $Var[\varepsilon_i^2] = 8\ m\ (m+2)(m+3)/m^4$. Therefore,

$$\begin{aligned} Var[\hat{\sigma}_{n,t}^4] &= \frac{m^2}{n^2} \sum_{i=1}^{m/n} Var[\sigma_{n,t,[i]}^4] + 2\frac{8m(m+3)}{n^2(m+2)} \sum_{i=1}^{m/n} E[\sigma_{n,t,[i]}^8] \\ &= Var[\sigma_{n,t}^4] + 2\frac{8m(m+3)}{n^2(m+2)} \sum_{i=1}^{m/n} E[\sigma_{n,t,[i]}^8] \end{aligned} \tag{A.4}$$

and hence, $\psi = \psi_0 / \left[1 + \frac{16}{n^2} \frac{m(m+3)}{m+2} \frac{\sum_{i=1}^{m/n} E[\sigma_{n,t,[i]}^8]}{Var(\sigma^4)}\right]$.

B Supplementary Material

Eric Ghysels, Per Mykland and Eric Renault (2021), *In-sample Asymptotics and Across-sample Efficiency Gains for High Frequency Data Statistics*, Econometric Theory Supplementary Material. To view, please visit: `https://doi.org/XXXXXX`

References

Aït-Sahalia, Y., J. Fan, R. J. Laeven, C. D. Wang, and X. Yang (2017): "Estimation of the continuous and discontinuous leverage effects," *Journal of the American Statistical Association*, 112, 1744–1758.

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., I. Kalnina, and D. Xiu (2020): "High-frequency factor models and regressions," *Journal of Econometrics*, 216, 86–105.

Aldous, D. J., and G. K. Eagleson (1978): "On Mixing and Stability of Limit Theorems," *Annals of Probability*, 6, 325–331.

Andersen, T., L. Benzoni, and J. Lund (2002): "An Empirical Investigation of Continuous-Time Equity Return Models," *Journal of Finance*, 57, 1239–1284.

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

Andersen, T., T. Bollerslev, and N. Meddahi (2005): "Analytic Evaluation of Volatility Forecasts," *Econometrica*, 73, 279–296.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001): "The Distribution of Realized Stock Return Volatility," *Journal of Financial Economics*, 61, 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001): "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42–55.

Andreou, E., and E. Ghysels (2002): "Rolling-Sample Volatility Estimators: Some New Theoretical, Simulation and Empirical Results," *Journal of Business and Economic Statistics*, 20, 363–376.

Barndorff-Nielsen, O., S. Graversen, J. Jacod, M. Podolskij, and N. Shephard (2006a): "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.

——— (2006b): "Limit theorem for bipower variations in financial econometrics," *Econometric Theory*, 22, 677–719.

Barndorff-Nielsen, O. E., B. Nielsen, N. Shephard, and C. Ysusi (2004): "Measuring and Forecasting Financial Variability Using Realised Variance with and Without a Model," in *State Space and Unobserved Component Models: Theory and Applications - Proceedings of a Conference in Honour of James Durbin*, ed. by S. J. K. Andrew C. Harvey, and N. Shephard. Cambridge University Press.

Barndorff-Nielsen, O. E., and N. Shephard (2001): "Non-Gaussian Ornstein-Uhlenbeck-Based Models and some of their Uses in Financial Econometrics," *Journal of the Royal Statistical Society, Series B*, 63, 167–241.

——— (2002a): "Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society, B*, 64, 253–280.

——— (2002b): "Estimating Quadratic Variation Using Realised Variance," *Journal of Applied Econometrics*, 17, 457–477.

——— (2004a): "Econometric analysis of realised covariation: high frequency based covariance, regression and correlation in financial economics," *Econometrica*, 72, 885–925.

——— (2004b): "Power and bipower variation with stochastic volatility and jumps (with discussion)," *Journal of Financial Econometrics*, 2, 1–48.

——— (2006): "Econometrics of testing for jumps in financial economics using bipower variation," *Journal of Financial Econometrics*, 4, 1–30.

——— (2007): "Variation, jumps, market frictions and high frequency data in financial econometrics," in *Advances in Economics and Econometrics. Theory and Applications, Ninth World Congress*, ed. by R. Blundell, P. Torsten, and W. K. Newey. Econometric Society Monographs, Cambridge University Press.

Bobkoski, M. (1983): "Hypothesis Testing in Nonstationary Time Series," Ph.D. Thesis, University of Wisconsin.

Bollerslev, T., A. J. Patton, and R. Quaedvlieg (2016): "Exploiting the errors: A simple approach for improved volatility forecasting," *Journal of Econometrics*, 192, 1–18.

Chan, N., and C. Wei (1987): "Asymptotic inference for nearly nonstationary AR(1) processes," *Annals of Statistics*, 15, 1050–1063.

Chen, X., and E. Ghysels (2011): "News - Good or Bad - and the impact on volatility predictions over multiple horizons," *Review of Financial Studies*, 24, 46–81.

Chernov, M., A. R. Gallant, E. Ghysels, and G. Tauchen (2002): "Alternative Models for Stock Price Dynamics," *Journal of Econometrics*, 116, 225–257.

Clark, T. E., and M. W. McCracken (2009): "Combining Forecasts from Nested Models," *Oxford Bulletin of Economics and Statistics*, 71, 303–329.

Comte, F., and E. Renault (1998): "Long memory in continuous-time stochastic volatility models," *Mathematical Finance*, 8, 291–323.

Elliott, G., T. Rothenberg, and J. Stock (1996): "Efficient tests for an autoregressive unit root," *Econometrica*, 64, 813–836.

Engle, R. F., and V. K. Ng (1993): "Measuring and testing the impact of news on volatility," *Journal of Finance*, 48, 1749–1778.

Foster, D. P., and D. B. Nelson (1996): "Continuous Record Asymptotics for Rolling Sample Variance Estimators," *Econometrica*, 64, 139–174.

Glynn, P., and D. L. Iglehart (1989): "The optimal linear combination of control variates in the presence of asymptotically negligible bias," *Naval Research Logistics*, 36, 683–692.

Gonçalves, S., and N. Meddahi (2009): "Bootstrapping realized volatility," *Econometrica*, 77, 283–306.

Hall, P., and C. C. Heyde (1980): *Martingale Limit Theory and Its Application*. Academic Press, Boston.

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

Jacod, J. (1994): "Limit of random measures associated with the increments of a Brownian semimartingale," Preprint number 120, Laboratoire de Probabilités, Université Pierre et Marie Curie, Paris.

——— (1996): "La variation quadratique du Brownian en presence d'erreurs d'arrondi," *Asterisque*, 236, 155–162.

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 A. N. Shiryaev (2003): *Limit Theorems for Stochastic Processes*. Springer-Verlag, New York.

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

——— (2014): "Efficient estimation of integrated volatility in presence of infinite variation jumps," *Annals of Statistics*, 42, 1029–1069.

——— (2018): "Limit theorems for integrated local empirical characteristic exponents from noisy high-frequency data with application to volatility and jump activity estimation," *Annals of Applied Probability*, 28, 511–576.

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. (2011): "Subsampling High Frequency Data," *Journal of Econometrics*, 161, 262–283.

Kalnina, I., and O. Linton (2007): "Inference about Realized Volatility using Infill Subsampling," LSE STICERD Research Paper No. EM523.

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.

Meddahi, N. (2002): "A theoretical comparison between integrated and realized volatility," *Journal of Applied Econometrics*, 17, 479–508.

Mykland, P. A., N. Shephard, and K. Sheppard (2012): "Efficient and feasible inference for the components of financial variation using blocked multipower variation," Working Paper, University of Oxford.

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

——— (2008): "Inference for Volatility Type Objects and Implications for Hedging," *Statistics and its Interface*, 1, 255–278.

——— (2009): "Inference for continuous semimartingales observed at high frequency," *Econometrica*, 77, 1403–1445.

——— (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.

——— (2017): "Assessment of Uncertainty in High Frequency Data: The Observed Asymptotic Variance," *Econometrica*, 85, 197–231.

Phillips, P. C. B. (1987): "Towards a unified asymptotic theory for autoregression," *Biometrika*, 74(3), 535–547.

Phillips, P. C. B., H. Moon, and Z. Xiao (2001): "How to estimate autoregressive roots near unity," *Econometric Theory*, 17, 29–69.

Phillips, P. C. B., and V. Solo (1992): "Asymptotics for linear processes," *Annals of Statistics*, pp. 971–1001.

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

Rényi, A. (1963): "On Stable Sequences of Events," *Sankyā Series A*, 25, 293–302.

Todorov, V., and G. Tauchen (2011): "Volatility jumps," *Journal of Business and Economic Statistics*, 29, 356–371.

Vetter, M. (2015): "Estimation of Integrated Volatility of Volatility with Applications to Goodness-of-fit Testing," *Bernoulli*, 21, 2393–2418.

Wang, C. D., and P. A. Mykland (2014): "The estimation of leverage effect with high-frequency data," *Journal of the American Statistical Association*, 109(505), 197–215.

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

Zhang, L. (2007): "What you don't know cannot hurt you: On the detection of small jumps," Discussion paper, University of Illinois at Chicago.

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.

Table 2: MSE improvements and weights for GARCH, Two Factor and Jump Diffusion Models

The design is detailed in section 5. The data generating processes appear in equations (5.1) through (5.4). More specifically, they are respectively referred to as a GARCH diffusion, two-factor SV diffusion and Jump diffusion. For the former two, the entries pertain to QV, with $(\omega^*_{ut})^{-1}$ appearing in equation (4.2), $(\omega^*_{vt})^{-1}$ appearing in equation (4.1) and $(\omega^{hc}_t)^{-1}$ taken from equation (4.8), whereas $(\omega^{unc})^{-1}$ is taken from equation (4.6). For the data generating process involving jumps the entries pertain to BPV and the weighting schemes appear in Table 1. The entries in the top panels are the ratios of MSE of the various estimators vis-à-vis the MSE of a standard RV/BPV estimator using the same frequency data. The lower panels report statistics of the weighting schemes (inverse to be more precise).

	GARCH Diffusion					
	ω^*_{ut}	ω^*_{vt}	ω^{hc}	ω^{unc}	ω^{rm}	ω^{r-th}
Mean MSE Impr.						
5 Min	0.87	0.85	0.95	0.95	0.86	0.87
10 Min	0.82	0.82	0.93	0.93	0.83	0.83
ω						
5 Min						
Mean	0.04	0.05	0.17	0.01	-	-
Ave. Var	0.02	0.01	0.01	0.00	-	-
10 Min						
Mean	0.07	0.07	0.23	0.01	-	-
Ave. Var	0.03	0.01	0.03	0.00	-	-
	Two-factor SV model					
Mean MSE Impr.						
5 Min	0.91	0.91	0.96	0.96	0.88	1.13
10 Min	0.87	0.87	0.94	0.94	0.85	1.19
ω						
5 Min						
Mean	0.03	0.02	0.14	0.11	-	-
Ave. Var	0.01	0.00	0.01	0.00	-	-
10 Min						
Mean	0.06	0.06	0.19	0.18	-	-
Ave. Var	0.01	0.00	0.03	0.02	-	-
	Jump Diffusion					
Mean MSE Impr.						
5 Min	0.99	0.99	0.99	0.99	0.89	0.94
10 Min	0.98	0.98	0.99	0.99	0.87	0.89
ω						
5 Min						
Mean	0.01	0.01	0.04	0.03	-	-
Ave. Var	0.00	0.00	0.00	0.01	-	-
10 Min						
Mean	0.01	0.01	0.08	0.07	-	-
Ave. Var	0.00	0.00	0.02	0.01	-	-

Table 3: MSE improvements and weights for Realized Betas

The design detailed in section 5 pertains to CAPM betas as in equation (3.21) with the market return process according to equation (5.2) with instantaneous volatility dynamics appearing in (5.4). The betas follow an Ornstein-Uhlenbeck process as in equation (5.6). We consider three DGPs. In particular, we take for DGP1 $\kappa = 2$, $\mu = 1.0$, with $\sigma_b = 3$, for DGP2 $\kappa = 2$, $\mu = 0.10$ and $\sigma_b = 0.03$, for DGP3 $\kappa = 1$, $\mu = 1.0$ and $\sigma_b = 0.03$. The weighting schemes are those appearing in Table 2. All entries pertain to 1 and 5 minute sampling schemes with sample size $T = 500$. The entries are the ratios of MSE of the various estimator vis-à-vis the MSE of a standard realized beta. In Panel B we report the mean of the weights.

		ω^*_{ut}	ω^*_{vt}	ω^{hc}	ω^{unc}	ω^{rm}
		Panel A: MSE improvements				
DGPs	Sampling					
1	1 min	0.82	0.75	0.80	0.80	0.83
	5 min	0.74	0.71	0.74	0.75	0.82
2	1 min	0.82	0.75	0.80	0.79	0.80
	5 min	0.77	0.71	0.75	0.76	0.81
3	1 min	0.81	0.75	0.80	0.80	0.83
	5 min	0.80	0.72	0.75	0.75	0.82
		Panel B: Mean weighting schemes				
1	1 min	0.16	0.11	0.12	0.13	-
	5 min	0.21	0.13	0.15	0.15	-
2	1 min	0.17	0.11	0.11	0.12	-
	5 min	0.20	0.13	0.14	0.15	-
3	1 min	0.18	0.10	0.11	0.11	-
	5 min	0.20	0.13	0.14	0.15	-