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

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

First Draft: December 2006
This Draft: January 6, 2010

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. Our analysis is reminiscent of local-to-unity asymptotics. 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 Fangfang Wang for excellent research assistance, participants at the 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 as well as Evan Anderson, Tim Bollerslev, Frank Diebold, Rob Engle, Nour Meddahi, Neil Shephard and Kevin Sheppard for comments.

†University of North Carolina, Department of Economics, Gardner Hall CB 3305, Chapel Hill, NC 27599-3305, phone: (919) 966-5325, e-mail: eghysels@unc.edu.

‡Oxford University and University of Chicago, Department of Statistics, 5734 University Avenue, Chicago, Illinois 60637, phone (773) 702-8044, e-mail: mykland@galton.uchicago.edu.

§University of North Carolina, Department of Economics, Gardner Hall CB 3305, Chapel Hill, NC 27599-3305, phone: (919) 966-5326, e-mail: renault@email.unc.edu.

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 (2002) as well as the recent survey by Barndorff-Nielsen and Shephard (2007).[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. This literature is surveyed by Andersen, Bollerslev, and Diebold (2009). The time series models that sit on top of the realized measures exploit the persistence properties of volatility, well documented in the prior literature on ARCH and Stochastic Volatility (SV) models (see Bollerslev, Engle, and Nelson (1994), Ghysels, Harvey, and Renault (1996), and Shephard (2004) for further references and details).

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. Take for example the measure of quadratic variation which has been the most widely studied. While in the limit in-sample observations suffice to estimate current realized variation, there are efficiency gains for any finite sample configuration, that is, there are gains to be made in practical applications of extracting realized volatility to use realized volatility from the previous days. Currently, the first step of the two-step procedure is completely detached from modeling, which is the subject of the second step. It is argued in this paper

[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 the survey by Barndorff-Nielsen and Shephard (2007) for relevant references).

that measuring volatility is not necessarily confined to a single (daily) interval and prior observations are useful thanks to persistence in volatility. As documented by Bandi, Russell, and Yang (2008) in the context of option pricing, we argue that even from a forecasting standpoint, there is scope for optimizing the finite sample properties of alternative volatility estimators.

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. For example, it has recently been documented by Gonçalves and Meddahi (2008) that improvement of inference on realized volatility through Edgeworth expansions is impaired by the lack of reliable estimators for the cumulants. In addition, our analysis applies to relatively well posed problems, such as estimating volatility from high frequency data *and* is even more relevant when trying to estimate less well posed quantities such as the estimation of the leverage effect.

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. The reason why this distinction is important is because it is often argued that volatility forecasting models are reduced form models which combine filtering and prediction, and it is the combination that matters most. This argument applies only to fixed parameter models, which embed fixed filtering schemes. Our filtering is 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.

The paper is organized as follows. Section 2 discusses the main motivation for our approach structured around three themes (1) efficiency gains, (2) connection with Kalman filtering, and (3) improvements in forecasting. Section 3 derives all our results using a simple framework The purpose of this section is to start with a relatively simple example that contains the core ideas of our analysis. The example is stylized for illustrative purpose - yet as we will show in section 4 the example turns out to be surprisingly comprehensive. Section 4 covers the case of general jump diffusion models. The theory is asymptotic in nature in terms of sampling of intra-daily data as well as the properties of the data generating process across days. Section 5 deals with the estimation of leverage. Section 6 reports simulation evidence on the efficiency and forecasting gains in a controlled environment of stylized diffusion processes. Section 7 concludes the paper.

2 Why Across-sample Weighting?

There is an abundance of data generated by the trading of financial assets around the world. This data-rich environment of high frequency intra-daily observations has ignited a very exciting research agenda of statistical analysis related to stochastic processes, in particular volatility. The key motivation for our paper, is the observation that it is both possible and relevant to improve intra-daily volatility measurements by taking advantage of previous days' information. There are three subsections that describe the main motivation along three themes (1) efficiency gains, (2) connection with Kalman filtering, and (3) improvements in forecasting.

2.1 Efficiency gains

Let us start with explaining why it is relevant, as one may wonder why we need an accurate measurement of volatility based on past information. In particular, one may argue that the prime objective is to forecast future volatility and that the use of past information via a preliminary filtering procedure is redundant, since prediction models capture past measurements. First, we show in this paper that even for the purpose of forecasting volatility, it cannot hurt to improve its measurement through filtering. This is particularly relevant for option pricing which involves nonlinear forecasts of future volatility. However, our main motivation is that volatility measurement per se is important in its own right for many financial applications, such as for example trading execution of limit orders, option hedging, volatility timing for portfolio management, Value-at-Risk computations, beta estimation, specification testing such as detecting jumps, among others.

The question remains whether it is actually possible to improve volatility measurement with using prior days' information as it is often argued that arbitrary frequently observed intra-day data provide exact observation of volatility. There are at least two reasons why the use of past information helps.

First, the actual number of intra-daily observations is not infinite and so called microstructure market frictions may prevent us from sampling too frequently. Moreover, recent procedures (kernel, two time-scale, etc.) that use data sampled at the highest frequency despite the presence of microstructure noise, feature an effective number of observations smaller than actually recorded. We can take advantage of volatility persistence in this regard. An important novel feature of our analysis is that it has some commonality with the local-to-unity asymptotics of Bobkoski (1983), Chan and Wei (1987), Elliott, Rothenberg, and Stock (1996), Phillips, Moon, and Xiao (2001), among many others. The original local-to-unity asymptotics was used to better approximate the finite sample behavior of parameter estimates in autoregressive models with roots near the unit circle where neither the Dickey-Fuller asymptotics nor the standard normal asymptotics provide adequate descriptions of the finite sample properties of OLS estimators. Here local-to-unity asymptotics is used to improve finite sample estimates too, albeit in a context of in-sampling asymptotics in the spirit of

Phillips, Moon, and Xiao (2001). As emphasized by Phillips et al., one advantage of this approach is that local-to-unity behavior is displayed by data inside each block (intraday data for us) while the daily volatility process behaves like a stationary process.Here local-to-unity asymptotics is used to improve finite sample estimates too, albeit in a context of in-sampling asymptotics.

Second, we do in fact not necessarily have to rely on local-to-unity asymptotic arguments. For example, existing estimators for the model-free measurement of leverage effects using intra-daily data are not consistent albeit asymptotically unbiased, see e.g. Mykland and Zhang (2009). Namely, even with an infinite number of intra-daily observations, current estimators of leverage effects converge in distribution to a random variable centered at the true value. For such estimators our analysis is even more relevant as it allows one to reduce the variance of measurement by combining several days of measurement. This approach is justified by the fact that financial leverage ought to be a persistent time series process.

2.2 Connection with Kalman filtering

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

$$\begin{aligned} X_t^{(n)} &= Y_t + c_t^{(n)} Z_t^{(n)} \\ Z_t^{(n)} &\sim N(0,1) \end{aligned} \tag{2.1}$$

Traditional Kalman filtering applies to a Gaussian measurement equation like (2.1) augmented with a Gaussian transition equation:

$$\begin{aligned} Y_t &= \alpha Y_{t-1} + g + \sqrt{H_t} u_t \\ u_t &\sim N(0,1) \end{aligned} \tag{2.2}$$

Note that the Kalman filter considered here is slightly more general than traditional one, since we allow for conditional heteroskedasticity in equations (2.1) and (2.2). Conditioning on the values of the variables $c_t^{(n)}$ and H_t allows us to consider heteroskedasticity as predetermined and to use standard Bayes formula to obtain recursion formulas for the filtering distribution, that is the conditional normal distribution of Y_t given $(X_t^{(n)}, X_{t-1}^{(n)}, \ldots, X_1^{(n)})$. If $N(Y_t^*, K_t)$ stands for this conditional distribution (the superscript n is omitted for convenience), the Kalman recursion formulas (see e.g. Williams (1991), p. 168) can be written:

$$\begin{aligned} \frac{1}{K_t} &= \frac{1}{\alpha^2 K_{t-1} + H_t} + \frac{1}{c_t^2} \\ \frac{Y_t^*}{K_t} &= \frac{\alpha Y_{t-1}^* + g}{\alpha^2 K_{t-1} + H_t} + \frac{X_t}{c_t^2} \end{aligned} \tag{2.3}$$

Hence, the filtered values are defined as:

$$Y_t^* = (1 - \omega_t) X_t + \omega_t (\alpha Y_{t-1}^* + g) \tag{2.4}$$

with:

$$\omega_t = \frac{c_t^2}{c_t^2 + \alpha^2 K_{t-1} + H_t} \tag{2.5}$$

As will be discussed in detail later, our way to take advantage of past information to improve upon the measurement X_t is quite similar in spirit with Kalman-Bucy filter with however a couple of important differences. First, we do not deal with a Gaussian transition equation. The processes we consider, like realized variance or bi-power variation (defined later) are often constrained to be positive (and thus cannot be Gaussian). However, even more importantly, we want to remain model free. Therefore, instead of the model (2.2), we will simply refer to a linear projection of the variable Y_t on some of its lagged values. $Y_{t|t-1}$ will denote such an optimal linear predictor for a given number of lags. If for instance we decide to use only one lag, α and g will be defined as (maintaining a model-free setting):

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

Note that the structure of equation (2.1) is maintained but it in a model-free context, up to some regularity conditions ensuring its asymptotic validity.

Second, without a Gaussian transition equation we have to give up the Bayes formula which enables us to compute conditional expectations and variances like (2.3). Moreover, maintaining the model-free setting, we do not want to define a filtered value function of *all* past information $(X_t^{(n)}, X_{t-1}^{(n)}, ..., X_1^{(n)})$, but instead only for a fixed number of lags every single day t. For instance, in the case of one lag, our filtered value will be defined by:

$$Z_t = (1 - \omega_t^*)X_t + \omega_t^*(\alpha X_{t-1} + g) \tag{2.6}$$

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

$$\omega_t^* = \frac{c_t^2}{c_t^2 + \alpha^2 c_{t-1}^2 + H_t} \tag{2.7}$$

Note that our optimal weight (2.7), albeit similar to the Kalman one (2.5), is smaller since $c_t^2 > K_t$ (using equation (2.3)). Intuitively, our weighting schemes give less weight to past information since we summarize past information by the measured counterpart $\alpha X_{t-1} + g$ (with conditional variance $\alpha^2 c_{t-1}^2$) of the past forecast $Y_{t|t-1} = \alpha Y_{t-1} + g$, whereas the Kalman filter weights use the more accurate filtered counterpart $\alpha Y_{t-1}^* + g$ (i.e. with smaller conditional variance $\alpha^2 K_{t-1} < \alpha^2 c_{t-1}^2$).

2.3 Forecasting gains

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

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

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

Suppose now that we also consider past "improved" observations: $Z_s = (1-\omega)X_s + \omega Y_s^*$, where Y_s^* is an unbiased linear predictor of X_s.[2] Since Y_{t+1}^* is unbiased linear predictor of X_{t+1} :

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

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

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

It is natural to assume in addition that:

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

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

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

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

$$Var(Y_{t+1}^X - Y_{t+1}) - Var(Y_{t+1}^Z - Y_{t+1}) = \omega^2(Var(v_{t+1}^X) - Var(v_{t+1}^*)) \tag{2.8}$$

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

The above result pertains to a simple linear forecasting setting. We expect that improvements of measurement are going to be even more important nonlinear settings. The most prominent

[2]We consider here weights that are not time varying - unlike in the previous subsection. Hence, we are assessing here the impact on forecasting performance of fixed weights ω. Optimally chosen time varying weights should ensure at least a comparable forecasting performance. We will refer to the latter as conditional schemes, in contrast to unconditional ones that are also discussed in subsection 4.2.

example where the objective of interest is a nonlinear function of future volatility, is option pricing. The future path of volatility until some time to maturity of the derivative contract determines the current option price through a conditional expectation of a nonlinear payoff function. A simplified example is the model of Hull and White (2005) where the price of a European call option of time to maturity h and moneyness k equals:

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

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

3 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 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:

$$r^n_{t,j} = p_{t-(j-1)/n} - p_{t-j/n} \quad (3.1)$$

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 j^{th} 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 3.1 *Within a day (period) t, given a sequence $\sigma_{t,j}^2$, $j = 1, \ldots, n$, the return process in equation (3.1) is distributed independently Gaussian:*

$$r_{t,j}^n \sim N(0, \frac{1}{n}\sigma_{t,j}^2) \tag{3.2}$$

for all $j = 1, \ldots, n$.

For every period t the parameter of interest is:

$$\sigma_t^2 \equiv Var(\sum_{j=1}^{n}[r_{t,j}^n]) \equiv \frac{1}{n}\sum_{j=1}^{n}\sigma_{t,j}^2 \tag{3.3}$$

and consider the following ML estimators for each t :

$$\hat{\sigma}_t^2 = \sum_{j=1}^{n}[r_{t,j}^n]^2 \tag{3.4}$$

Then conditional on the volatility path $\sigma_{t,j}^2$, $j = 1, \ldots, n$, we have, under Assumption 3.1 the following properties for the ML estimators:

$$E_c[\hat{\sigma}_t^2] = \sigma_t^2 \tag{3.5}$$

$$Var_c[\hat{\sigma}_t^2] = \frac{2}{n^2}\sum_{j=1}^{n}\sigma_{t,j}^4 = \frac{2}{n}\sigma_t^{[4]} \tag{3.6}$$

where $\sigma_t^{[4]} = 1/n\sum_{j=1}^{n}\sigma_{t,j}^4$, $E_c[\cdot] = E[\cdot|\sigma_{t,j}^2, \forall j]$ and similarly for $Var_c[\cdot] = Var[\cdot|\sigma_{t,j}^2, \forall j]$.

In a first subsection 3.1 we address the key question of the paper in the context of the simple example, namely to what extend can we improve the estimation of σ_t^2 using prior day information. After characterizing the optimal weighting scheme we discuss its asymptotic relevance in subsection 3.2 and related estimation issues in subsection 3.3.

3.1 Characterizing the Optimal Weighting Scheme

The question we address is to what extend can we improve the estimation of σ_t^2 using prior day information, and in particular using $\hat{\sigma}_{t-1}^2$. We want to do this with a model-free approach, using optimal linear predictors in the spirit of (2.6). To formalize this, assume that:

Assumption 3.2 *The n-dimensional process* $(\sigma^2_{t,j})_{1\leq j\leq n}$, $t \in \mathbb{Z}$ *is weakly stationary.*

Assumption 3.2 implies that the estimated process $(\hat{\sigma}^2_t)$ is itself weakly stationary, with second order characteristics given by (see Appendix A for a proof):

$$\begin{aligned} E(\hat{\sigma}^2_t) &= E(\sigma^2_t) = \sigma^2 \\ Cov(\hat{\sigma}^2_t, \hat{\sigma}^2_{t-h}) &= Cov(\sigma^2_t, \sigma^2_{t-h}), h > 0 \\ Var(\hat{\sigma}^2_t) &= Var(\sigma^2_t) + \frac{2}{n} E(\sigma^{[4]}_t) \end{aligned} \tag{3.7}$$

Then, the optimal linear predictors of $\hat{\sigma}^2_t$ given $\hat{\sigma}^2_{t-1}$, will be written as:

$$\hat{\sigma}^2_{t|t-1} = (1-\varphi)\sigma^2 + \varphi\hat{\sigma}^2_{t-1} \tag{3.8}$$

where:

$$\varphi \equiv \frac{Cov(\hat{\sigma}^2_t, \hat{\sigma}^2_{t-1})}{Var(\hat{\sigma}^2_t)} \tag{3.9}$$

The goal is to combine $\hat{\sigma}^2_{t|t-1}$ and $\hat{\sigma}^2_t$, to improve the estimation of σ^2_t using prior day information. In particular, we define a new estimator combining linearly $\hat{\sigma}^2_{t|t-1}$ and $\hat{\sigma}^2_t$:

$$\begin{aligned} \hat{\sigma}^2_t(\omega_t) &= (1-\omega_t)\hat{\sigma}^2_t + \omega_t\hat{\sigma}^2_{t|t-1} \\ &= \hat{\sigma}^2_t - \omega_t(\hat{\sigma}^2_t - \hat{\sigma}^2_{t|t-1}) \end{aligned} \tag{3.10}$$

Note that the weight ω_t depends on t, as indeed it a conditional weighting scheme, and its computation will be volatility path dependent. To characterize the optimal weighting scheme, one may apply a conditional control variables principle, given the volatility path.

The optimal weighting scheme will be denoted ω^*_t. For notational simplicity, the conditioning is not made explicit in the formulas below and the criterion to minimize will be written as:

$$\omega^*_t \equiv \text{Argmin}_{\omega_t} E_c[\hat{\sigma}^2_t(\omega_t) - \sigma^2_t]^2 = \text{Argmin}_{\omega_t} E_c\{\hat{\sigma}^2_t - \sigma^2_t - \omega_t(\hat{\sigma}^2_t - \hat{\sigma}^2_{t|t-1})\}^2 \tag{3.11}$$

We will need to rely on a optimal control variables result to derive the optimal weighting scheme, in particular results that take into account the possibility of bias since we need to take into account the non-zero mean of $(\hat{\sigma}^2_t - \hat{\sigma}^2_{t|t-1})$ given the volatility path. Such a result was derived - see Proposition 1 in Glynn and Iglehart (1989) - and we state it as the following

Lemma:

Lemma 3.1 *If $\overline{\theta}$ is an unbiased estimator of θ, u a zero-mean random variable and c a given real number, an estimator of θ with minimum mean squared error in the class of estimators: $\overline{\theta}(\omega) = \overline{\theta} - \omega(u + c)$ is obtained as $\overline{\theta}(\omega^*)$ with: $\omega^* = Cov[\overline{\theta}, u]/(Var(u) + c^2)$.*

Applying Lemma 3.1 the ω_t^* is obtained as:

$$\omega_t^* = \frac{Cov_c[\hat{\sigma}_t^2, \hat{\sigma}_t^2 - \hat{\sigma}_{t|t-1}^2]}{Var_c(\hat{\sigma}_t^2 - \hat{\sigma}_{t|t-1}^2) + [E_c(\hat{\sigma}_t^2 - \hat{\sigma}_{t|t-1}^2)]^2} \tag{3.12}$$

Note that ω_t^* has been shrunk with respect to the regression coefficient of $\hat{\sigma}_t^2$ on ($\hat{\sigma}_t^2$ - $\hat{\sigma}_{t|t-1}^2$), and this is due - as noted before - to the need to take into account the non-zero mean of $(\hat{\sigma}_t^2 - \hat{\sigma}_{t|t-1}^2)$ given the volatility path.

From here we proceed with specific assumptions for the data generating process that will yield explicit expressions for the components of ω_t^* in equation (3.12). In the present section we have a fairly simple setting for the data generating process that will allow us to derive the optimal weighting scheme as follows:

Proposition 3.1 *Consider data generated as specified in Assumption 3.1. Then, the optimal weighting scheme equals:*

$$\omega_t^* = \frac{2\sigma_t^{[4]}}{2[\sigma_t^{[4]} + \varphi^2\sigma_{t-1}^{[4]}] + n[\sigma_t^2 - \varphi\sigma_{t-1}^2 - (1-\varphi)\sigma^2]^2} \tag{3.13}$$

Proof: See Appendix A.

The result of Proposition 3.1, albeit in a very stylized context, is sufficient to present the main motivation of our paper. The key issue is to assess to what extent our preferred modified estimator $\hat{\sigma}^2(\omega_t^*)$ is preferred to the MLE $\hat{\sigma}_t^2 \equiv \hat{\sigma}_t^2(0)$ based exclusively on day t intradaily observations. Hence, the question arises whether we want to pay the price of a bias proportional to the conditional prediction bias $E_c[\hat{\sigma}_t^2 - \hat{\sigma}_{t|t-1}^2]$ to lower the conditional variance because of the control variables principle. This is actually an empirical question. From Proposition 3.1, we see that, not surprisingly, ω_t^* goes to zero when n goes to infinity for a given non-zero value of the bias:

$$E_c[\hat{\sigma}_t^2 - \hat{\sigma}_{t|t-1}^2] = \sigma_t^2 - (1-\varphi)\sigma^2 - \varphi\sigma_{t-1}^2 \tag{3.14}$$

This may lead one to believe that $\hat{\sigma}_t^2(0)$ should be our preferred estimator. However, in practice, n is never infinitely large and the squared bias $[E_c[\hat{\sigma}_t^2 - \hat{\sigma}_{t|t-1}^2]]^2$ may be relatively small with respect to the gain in variance by control variables, which is approximately proportional to $[Var_c[\hat{\sigma}_t^2 - \hat{\sigma}_{t|t-1}^2]] = 2\sigma_t^{[4]}/n + 2\varphi^2\sigma_{t-1}^{[4]}/n$. This is precisely the bias-variance trade off that we can notice in the denominator of the above formula for the optimal weight ω_t^*.

Two remarks are in order. First of all, note that the optimal weight ω_t^* will automatically become arbitrarily close to zero whenever the bias 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 ω_t^* brings us back to standard MLE. Second, our Monte Carlo experiments will confirm that the optimal weight is not negligible in general. Its order of magnitude is rather between 10 % and 30 %. We can actually set down a formal framework to rationalize these findings. To do so, note that is worth relating the feasible bias:

$$B_t^F \equiv \sigma_t^2 - (1-\varphi)\sigma^2 - \varphi\sigma_{t-1}^2$$

with the infeasible bias:

$$B_t^U \equiv \sigma_t^2 - (1-\varphi_0)\sigma^2 - \varphi_0\sigma_{t-1}^2 \tag{3.15}$$

where φ_0 stands for the auto-correlation coefficient computed on the true unknown volatility process:

$$\varphi_0 = \frac{Cov(\sigma_t^2, \sigma_{t-1}^2)}{Var(\sigma_t^2)}$$

The difference B_t^U - $B_t^F = (\varphi - \varphi_0)(\sigma_{t-1}^2 - \sigma^2)$ will be of order $1/n$ because of the following result:

Proposition 3.2 *Consider data generated as specified in Assumption 3.1. Then:*

$$\varphi_0 - \varphi = (2\varphi/n)E(\sigma_t^{[4]})[Var(\sigma_t^2)]^{-1} \tag{3.16}$$

Moreover, the size of the theoretical bias B_t^U is tightly related to volatility persistence. More *precisely, we can write:*

$$B_t^U \equiv \sigma_t^2 - (1-\varphi_0)\sigma^2 - \varphi_0\sigma_{t-1}^2 = [1-\varphi_0^2]^{1/2}U_t$$

where the process U_t has unconditional mean zero and unconditional variance equal to $[Var(\sigma_t^2)]$.

Proof: See Appendix A.

Since:

$$B_t^F = B_t^U + O(1/n)$$

it is equivalent to show that the feasible squared bias bias or the unfeasible one will not necessarily dominate the conditional standard error $2\sigma_t^{[4]}/n + 2\varphi^2\sigma_{t-1}^{[4]}/n$. In other words, following proposition 3.1, we would like to show that:

$$1 - \varphi_0^2 = O(1/n)$$

To do so, the logic of our approach rests upon the new framework of near integration as formulated by Phillips, Moon, and Xiao (2001) (PMX hereafter).

3.2 A block local to unity framework for volatility

The goal of this subsection 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 as a function of n, hence $\varphi_0(n)$, such that:

$$n[1-\varphi_0^2(n)] = O(1) \tag{3.17}$$

One way to achieve (3.17) is to assume:

$$\varphi_0(n) = 1 - (c/n) \qquad c > 0 \tag{3.18}$$

Hence, we have a drifting Data Generating Process ($\varphi_0(n)$ is increasing with n) to capture the notion that as n increases, we require more persistence in the intraday volatility process to ensure that the forecast $\hat{\sigma}_{t|t-1}^2$ of σ_t^2 - which uses past information - still improves $\hat{\sigma}_t^2$, which uses n intra-daily observations. To justify an assumption like (3.18), we propose to see the intraday volatility process as an autoregressive of order one process:

$$\sigma_{t,j}^2 = (1-\rho)\sigma^2 + \rho\sigma_{t,j-1}^2 + u_{t,j}, 1 < j \leq n \tag{3.19}$$

where $(u_{t,j})_{1\leq j\leq n}$ is a weak white noise. Note that equation (3.19) is only an example of model we consider to justify our assumption in equation (3.17) and is not a necessary condition. Note also that since, in our setting, the time interval Δt between two consecutive intraday observations is varying with n, standard formulas of temporal aggregation of autoregressive processes of order 1 lead us to consider the autocorrelation coefficient like $\rho = \exp(-k\Delta t)$ for some $k > 0$. With the convention of the previous subsection that one day corresponds to one unit of time, we will have $\Delta t = (1/n)$ but, since PMX do not make explicit any relationship between the size n of a block and the time interval between consecutive observations within a block (they keep $\Delta t = 1$), it is better to adapt their setting with the general notation Δt. Like PMX, we assume that the autoregression (3.19) is the same within each block (each day for us) and that the initial conditions in each block are set so that they correspond to the final observation in the previous block:

$$\sigma_{t,1}^2 = (1-\rho)\sigma^2 + \rho\sigma_{t-1,n}^2 + u_{t,1}$$

With this setting, it is easy to compute the population correlation coefficient φ_0. For large n, we have (see Appendix A):

$$\varphi_0 \sim \frac{[1-\exp(-k)]^2}{2[\exp(-k)-1+k]} \tag{3.20}$$

Following PMX, we now consider local-to-unity asymptotics in each block by localizing in the following way:

$$k \sim \frac{d}{n}, d > 0 \tag{3.21}$$

Note that in the context of PMX ($\Delta t = 1$), (3.21) exactly corresponds to their block-local-to-unity assumption ($\rho = 1 - \frac{d}{n}$). Plugging (3.21) into (3.20), we deduce:

$$\varphi_0 \sim 1 - \frac{d}{6n}$$

Therefore, our maintained assumption appearing in equation (3.17), and its particular specification in (3.18) as well are both 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), 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 (3.18) is governed by n, i.e. the number of intradaily data. In this respect, what is really required for our approach is in fact:

$$[1-\varphi_0^2(n)] = O(\varphi_0(n)-\varphi(n)) \tag{3.22}$$

where the notation $O(.)$ must be understood as an upper bound. Note that $[1-\varphi_0^2(n)]$ and $(\varphi_0(n)-\varphi(n))$ 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 $(\varphi_0(n)-\varphi(n))$ 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 (3.22) is that, roughly speaking, the level of volatility persistence is as least as good as the quality of our intradaily volatility estimator. It ensures that the squared feasible bias:

$$(B_t^F)^2 = O([1-\varphi_0^2(n)]) = O(\varphi_0(n)-\varphi(n)) \tag{3.23}$$

does not dominates the conditional variance $Var_c(\hat{\sigma}_t^2 - \hat{\sigma}_{t|t-1}^2)$. Note also that for given n, the time series (σ_t^2) and $(\hat{\sigma}_t^2)$ are still stationary processes, albeit highly persistent when n is large. In particular, a long time series $(\hat{\sigma}_t^2), t = 1, ..., T$ allows to consistently estimate (at standard rate $\sqrt{T}$) the unconditional mean σ^2 and the correlation coefficient φ. We will always consider that the time span is large enough to make $(1/\sqrt{T})$ small compared to $(1/n)$ (or at least $O(\varphi_0(n)-\varphi(n))$) in order to make negligible the time series estimation error in the parameters σ^2 and φ. Interestingly enough, Theorem 2 of PMX shows that the limit value $\sqrt{T} = n$ would restore a rate of convergence equal to T, as in the standard unit root setting, for the OLS estimation of φ_0 based on the complete set of hypothetical observations $\sigma_{t,j}^2$. For all practical purpose, our Monte Carlo experiments have confirmed that the time series estimation error on the parameters σ^2 and φ is negligible in front of $O(\varphi_0(n)-\varphi(n))$.

3.3 Estimating Optimal Weights

Having characterized the optimal weighting scheme we now turn to estimation issues. From Proposition 3.1 we know that the optimal weighting scheme ω_t^* depends on $\sigma_t^{[4]}$ and φ. 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.

We first turn to the MLE estimation of $Var\hat{\sigma}_t^2 = 2\sigma_t^{[4]}/n$ where:

$$\sigma_t^{[4]} = \frac{1}{n}\sum_{j=1}^{n}\sigma_{t,j}^4$$

To proceed we will make the following assumption:

Assumption 3.3 *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 3.3 the MLE $\hat{\sigma}_{t,[i]}^2$ of $\sigma_{t,[i]}^2$ is:

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

Then the MLE of $\sigma_{t,[i]}^4$ is such that:

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

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

$$\begin{aligned} \hat{\sigma}_t^{[4]} &= \frac{m}{n}\sum_{i=1}^{n/m}\frac{\hat{\sigma}_{t,[i]}^4}{1+2/m} \\ &= \frac{n}{m+2}\sum_{i=1}^{n/m}\left[\sum_{j=m(i-1)+1}^{mi} r_{t,j}^2\right]^2 \end{aligned} \tag{3.24}$$

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

$$\begin{aligned}
\tilde{\sigma}_t^{[4]} &= \frac{n}{3}\sum_{j=1}^{n} r_{t,j}^4 \\
&= \frac{n}{3}\sum_i \sigma_{t,[i]}^4 \sum_j (\frac{r_{t,j}}{\sigma_{t,[i]}})^4 \\
&\backsim \frac{1}{3n}\sum_i \sigma_{t,[i]}^4 \sum_j (\chi^2(1))^2
\end{aligned} \tag{3.25}$$

In Appendix B we compare the efficiency of the estimators $\hat{\sigma}_t^4$ and $\tilde{\sigma}_t^4$, showing that when m > 1, the former will be more efficient.

Recall that we try to improve the estimation of σ_t^2 using prior day information, and in particular using $\hat{\sigma}_{t-1}^2$. This argument is not confined to volatility measures. In particular, we can use the arguments spelled out so far to improve upon $\hat{\sigma}_t^{[4]}$ by using estimates from prior observation intervals. Namely, we can also consider the correlation coefficient ψ of $\hat{\sigma}_t^{[4]}$ on $\hat{\sigma}_{t-1}^{[4]}$:

$$\psi = \frac{Cov(\hat{\sigma}_t^{[4]}, \hat{\sigma}_{t-1}^{[4]})}{Var(\hat{\sigma}_t^{[4]})} \tag{3.26}$$

and develop a filtering procedure for $\sigma_t^{[4]}$ similar to the one proposed above for σ_t^2. Explicit formulas are provided in Appendix B.

4 General Theory

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. To do so, we consider continuous time jump diffusion models. The theory here is asymptotic in nature in terms of sampling of intra-daily data as well as the properties of the data generating process across days. More specifically, we assume a continuous time stochastic volatility jump-diffusion model for asset returns, namely:

$$dp\,(t) = \mu\,(t)\,dt + \sigma\,(t)\,dW\,(t) + \kappa\,(t)\,dq\,(t) \tag{4.1}$$

where $dq(t)$ is a counting process with $dq(t) = 1$ corresponding to a jump at t and $dq(t) = 0$ if no jump. The (possibly time-varying) jump intensity is $\lambda(t)$ and $\kappa(t)$ is the jump size. We are interested in measures such as the increments of quadratic variation:

$$QV_t = \sigma_t^{[2]} + \sum_{\{s \in [t-1,t]: dq(s)=1\}} \kappa^2(s). \tag{4.2}$$

where $\sigma_t^{[2]} = \int_{t-1}^{t} \sigma^2(s)\, ds$ corresponding to the continuous path component.

To make the explicit connection between the discrete time processes in Section 3 and the continuous time setting here we refer the reader to Appendix C, where we provide the steps required, to make the transition. Essentially, it is shown in Appendix C that the development in Section 3 remains valid so long as all variances and MSEs are interpreted asymptotically. Regarding the data generating process, we rely again on a local-to-unity asymptotic argument which consists of assuming that the persistence across days is sufficiently high to achieve an asymptotic trade-off with the intra-daily sampling frequency going to infinity. The first subsection 4.1 covers the estimation of quadratic variation without jumps followed by a subsection 4.2 distinguishing conditional weighting schemes discussed so far and unconditional schemes suggested in some of the prior literature. Subsection 4.3 covers bi-power variation and quarticity, that is statistics measuring quadratic variation in the presence of jumps and high order moments, as well as the issue of microstructure noise and leverage.

4.1 The case of Quadratic Variation without Jumps

We consider the estimation of quadratic variation (or more specifically integrated volatility in this subsection) for model (4.1), and for simplicity we will focus first on the case without jumps. There is now a well established literature on the estimation and usage of such measures. The volatility measures appearing in equation (4.2) are not observable but can be estimated from data. It is possible to consistently estimate QV_t in (4.2) by summing squared intra-daily returns, yielding the so called realized variance, namely:

$$\overline{QV}_t^n = \sum_{j=1}^{n} \left(r_{t,j}^n\right)^2. \tag{4.3}$$

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.

$$\lim_{n \to \infty} \overline{QV}_t^n \to^p QV_t. \tag{4.4}$$

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

$$\frac{n^{\alpha/2}(\overline{QV}_t - QV_t)}{\sqrt{2Q_t}} \xrightarrow{d} N(0,1) \tag{4.5}$$

where $Q_t = \int_{t-1}^t \sigma(s)^4 ds$ is called the quarticity. Note that for purpose of later extensions we parameterize the rates of convergence with α. In the current subsection we set $\alpha = 1$. It should be noted that in the case of no leverage effect, the result in (4.5) follows directly from the simplified example in Section 3 - in combination with the discussion in Appendix C. The case with leverage will be discussed later in 4.3.3.

We want to estimate $QV_t = \int_{t-1}^t \sigma_s^2 ds$ and take advantage of observations on the previous day, summarized by $\overline{QV}_{t-1}$, the estimator of QV_{t-1}. In general, we could take advantage of more than one lag, i.e. $\overline{QV}_{t-i}$ for $i > 1$, but for the moment we simplify the exposition to a Markov process of order one. The key assumption is that the two estimators $\overline{QV}_i$, $i = t-1$ and t have an asymptotic accuracy of the same order of magnitude and are asymptotically independent, for a given volatility path, namely:

$$\begin{aligned} \frac{n^{\alpha/2}(\overline{QV}_{t-1} - QV_{t-1})}{\sqrt{2Q_{t-1}}} &\xrightarrow{d} N(0,1) \\ \frac{n^{\alpha/2}(\overline{QV}_t - QV_t)}{\sqrt{2Q_t}} &\xrightarrow{d} N(0,1) \end{aligned} \tag{4.6}$$

and the joint asymptotic distribution is the product of the marginals.[3] We consider possible improvements of our estimator of QV_t, assuming for the moment that we know the correlation coefficient:

$$\varphi = \frac{Cov(\overline{QV}_t, \overline{QV}_{t-1})}{Var(\overline{QV}_{t-1})} \tag{4.7}$$

and the unconditional expectation $E(\overline{QV}_t) = E(\overline{QV}_{t-1}) = E(QV)$. Note that equation (4.7)

[3] Later in the paper we will consider more general settings where the asymptotic variance is not as simple. Throughout our analysis we will maintain the assumption that all randomness in the asymptotic variance goes through the volatility paths $(\sigma_s^2)_{s \in [t-2,t-1]}$ and $(\sigma_s^2)_{s \in [t-1,t]}$.

does *not* imply that our analysis is confined to AR(1) models. Instead, equation (4.7) only reflects the fact that we condition predictions on a single lag $\overline{QV}_{t-1}$. There may be potential gains from considering more lags, as the underlying models would result in higher order dynamics. Yet, for our analysis we currently focus exclusively on prediction equations with a single lag. Higher order equations are a straightforward extension discussed later.

The theory presented in the sequel will mirror the development in Section 3, but be valid even when volatility is not piecewise constant.

Consider the best linear forecast of $\overline{QV}_t$ using (only) $\overline{QV}_{t-1}$:

$$\overline{QV}_{t|t-1} = \varphi \overline{QV}_{t-1} + (1-\varphi)E(QV) \tag{4.8}$$

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

Our goal is to combine the two measurements $\overline{QV}_t$ and $\overline{QV}_{t|t-1}$ of QV_t to define a new estimator:

$$\overline{QV}_t(\omega_t) = (1-\omega_t)\overline{QV}_t + \omega_t \overline{QV}_{t|t-1} \tag{4.9}$$

Intuitively, the more persistent the volatility process, the more $\overline{QV}_{t|t-1}$ is informative about $\overline{QV}_t$ and the larger the optimal weight ω_t should be. Note that the weight depends on t, as indeed its computation will be volatility path dependent. To characterize such an optimal choice, one may apply a conditional control variable principle, given the volatility path. For notational simplicity, the conditioning is not made explicit in the formulas below and the criterion to minimize will be written as:

$$E[\overline{QV}_t(\omega_t) - QV_t]^2 = E\{\overline{QV}_t - QV_t - \omega_t(\overline{QV}_t - \overline{QV}_{t|t-1})\}^2 \tag{4.10}$$

Then, it can be shown that:

$$\overline{QV}_t(\omega_t^*) = \overline{QV}_t - \omega_t^*(\overline{QV}_t - \overline{QV}_{t|t-1}) \tag{4.11}$$

will be an optimal improvement of $\overline{QV}_t$ if ω_t^* is defined according to the following control variable formula:

$$\omega_t^* = \frac{Cov[\overline{QV}_t, \overline{QV}_t - \overline{QV}_{t|t-1}]}{Var(\overline{QV}_t - \overline{QV}_{t|t-1}) + [E(\overline{QV}_t - \overline{QV}_{t|t-1})]^2} \tag{4.12}$$

Note that ω_t^* has been shrunk with respect to the regression coefficient of $\overline{QV}_t$ on ($\overline{QV}_t$ - $\overline{QV}_{t|t-1}$). This is due to the need to take into account the non-zero mean of $(\overline{QV}_t - \overline{QV}_{t|t-1})$ given the volatility path.

A closed form formula for the optimal weights ω_t^* is obtained by computing moments, given the volatility path, according to the asymptotic distribution appearing in (4.6). Then, given the volatility path, we have:

$$\begin{aligned}
E(\overline{QV}_t - \overline{QV}_{t|t-1}) &= [QV_t - \varphi QV_{t-1} - (1-\varphi)E[QV_t]] \\
Var(\overline{QV}_t - \overline{QV}_{t|t-1}) &= \frac{2Q_t}{n^\alpha} + \varphi^2 \frac{2Q_{t-1}}{n^\alpha} + o(\frac{1}{n^\alpha}) \\
Cov[\overline{QV}_t, \overline{QV}_t - \overline{QV}_{t|t-1}) &= \frac{2Q_t}{n^\alpha} + o(\frac{1}{n^\alpha})
\end{aligned}$$

Therefore, ω_t^* defined by (4.12) can be rewritten as:

$$\omega_t^* = \frac{2Q_t}{2Q_t + \varphi^2 2Q_{t-1} + n^\alpha (B_t^F)^2} + o(\frac{1}{n^\alpha})$$

where $B_t^F = QV_t$ - φQV_{t-1} - $(1-\varphi)E[QV]$. Note that for the case of realized volatility, equation (4.12) is a corollary to Proposition 3.1, in view of the discussion in Appendix C.

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 $2Q_t/n^\alpha$ of $\overline{QV}_t$ is large in comparison to both (1) the asymptotic variance $2Q_{t-1}/n^\alpha$ of $\overline{QV}_{t-1}$ as well as (2) the quadratic forecast error $(B_t^F)^2$. However, for a given non-zero forecast error, the optimal weight ω_t^* goes to 0 when n goes to infinity. The reason is fairly straightforward: since $\overline{QV}_t$ is a consistent estimator of QV_t, forecasting QV_t from QV_{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 $2Q_t$ and the forecast error $(B_t^F)^2$. To correctly assess the latter, it is worth noting that:

$$B_t^F = B_t^F - B_t^U + B_t^U = (\varphi_0 - \varphi)(QV_{t-1} - E[QV_t]) + (1-\varphi_0)^{1/2}U_t \tag{4.13}$$

where, by a simple argument of variance decomposition, the variable U_t has a zero unconditional mean and an unconditional variance equal to $Var(QV_t) = Var(QV_{t-1}) = Var(U_t)$. The relevant trade-off is then clearly captured by the product $n^\alpha(1-\varphi_0^2)$, and therefore daily integrated volatility needs to be sufficiently persistent (φ_0 sufficiently close to 1) in comparison of the effective number n^α of intraday observations.

To proceed with the formal analysis, it will be convenient to make the persistence a function of n, hence $\varphi_0(n)$. More precisely, and in analogy with equation (3.17), let us assume that for some given number γ we have:

$$n^\alpha(1-\varphi_0(n)^2) = \gamma^2 \tag{4.14}$$

Hence, as discussed in subsection 3.2, similar to PMX we have a drifting Data Generating Process ($\varphi_0(n)$ increasing with n) to capture the idea that, the larger n is, the larger volatility persistence $\varphi_0(n)$ must be, to ensure that using the forecast $QV_{t|t-1}$of QV_t from QV_{t-1} still improves our estimator $\overline{QV}_t$ based on n intraday data. On the other hand, a straightforward extension of the proof of Proposition 3.2 gives $(\varphi_0 - \varphi) = O(\frac{1}{n^\alpha})$. Then the optimal weight is:

$$\omega_t^* = \frac{Q_t}{Q_t + (1-\frac{\gamma^2}{n^\alpha})Q_{t-1} + \gamma^2 U_t^2/2} + o(\frac{1}{n^\alpha}) = \frac{Q_t}{Q_t + Q_{t-1} + \gamma^2 U_t^2/2} + O(\frac{1}{n^\alpha}) \tag{4.15}$$

For large n, ω_t^* is, as expected, a decreasing function of γ^2. Larger the volatility persistence φ, smaller γ^2 and larger the weight ω_t^* assigned to day $t-1$ realized volatility to achieve day t improved volatility estimation.

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

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 Q_t with a sample equivalent, namely, $RQ_t = \sum_j^n \left(r_{t,j}^n\right)^4$. In the presence of jumps one needs to use tri- or quad-power variation, defined as: $TQ_{t+1,n} = n\mu_{4/3}^{-3}\Sigma_{j=3}^n \left|r_{t,j}\right|^{4/3} \left|r_{t,(j-1)}\right|^{4/3} \left|r_{t,(j-2)}\right|^{4/3}$, where $\mu_{4/3}$= $2^{2/3}\Gamma\left(7/6\right)\Gamma\left(0.5\right)^{-1}$. Along similar lines, the feasible asymptotic limit therefore implies that $\overline{BPV}_t$- $\int_{t-1}^t \sigma^2\left(s\right)ds \sim N\left(0, 0.6090RQ_t\right)$. These sample equivalents of quarticity will have to be used in the determination of the optimal weights. Improved estimation of quarticity can be derived in the spirit of section 3.3 (see Mykland and Zhang (2009)).

Besides the estimation of quarticity we face another problem. Namely, consider equation (4.14) and the resulting weighting scheme appearing in (4.15). Since $Var(QV_t) = Var(U_t)$, we can use a Kalman filtering like approximation to rewrite equation (4.15) in terms of inverse of optimal weights as:

$$
\begin{aligned}
[\omega_t^*]^{-1} &= 2 + \frac{Q_{t-1} - Q_t}{Q_t} + \gamma^2 \frac{Var[QV_t]}{2Q_t} \\
&= 2 + \frac{Q_{t-1} - Q_t}{Q_t} + \gamma^2 \frac{Var[U_t]}{2Q_t}
\end{aligned} \tag{4.16}
$$

In practice the term $\gamma^2 Var[QV_t]/Q_t$ will be computed as the ratio of $(1-\varphi^2)Var[QV_t]$ and the asymptotic (conditional) variance Q_t/n^α of the estimation error on integrated volatility.

An alternative approach to obtain a feasible estimator is to consider another expression for $(1-\varphi_0^2)^{1/2}U_t$ in equation (4.15). First, we should note that we do not observe QV_t - $QV_{t|t-1} = (1-\varphi_0^2)^{1/2}U_t$ but instead $\overline{QV}_t$ - $\overline{QV}_{t|t-1}$ which may differ from the true error by an estimation error of order $O(1/\sqrt{n^\alpha})$. However, such an error is of the same order as the object of interest $(1-\varphi_0^2)^{1/2}U_t$. Therefore, as a proxy we may consider the following feasible estimator:

$$
\begin{aligned}
[\omega_t^*]^{-1} &= 2 + \frac{Q_{t-1} - Q_t}{Q_t} + \gamma^2 \frac{(\overline{QV}_t - \overline{QV}_{t|t-1})^2}{2(1-\varphi^2)Q_t} \\
&= 2 + \frac{Q_{t-1} - Q_t}{Q_t} + \frac{(\overline{QV}_t - \overline{QV}_{t|t-1})^2}{2Q_t/n^\alpha}
\end{aligned} \tag{4.17}
$$

The latter will be useful to explore alternative weighting schemes, a topic to which we turn in the next subsection.

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 ADF representation to accommodate the local-to-unity asymptotics, following Stock (1991):

$$\Delta QV_t = c + \varphi_0 QV_{t-1} + \sum_{i=1}^{p-1} \varphi_{0i} \Delta QV_{t-i} + \varepsilon_t \tag{4.18}$$

where φ_0 is the sum of the autoregressive coefficients φ_{0i}, $i = 1, ..., p$. Following Stock (1991) we can apply local-to-unity asymptotics to the sum of AR coefficients, i.e. make $\varphi_0(n)$ a function of n, as we did in the AR(1) case. We leave this for future research.

4.2 Alternative Weighting Schemes

What sets this paper apart from previous attempts is the introduction of conditional information. The literature prior to our work consists of two contributions, Andreou and Ghysels (2002) and the unpublished section of Meddahi (2002). Both used unconditional adjustments, that is corrected volatility measure via time invariant schemes. The purpose of this section is to shed light on the advantages of using conditional information. We accomplish this goal by walking step-by-step from the unconditional to the optimal model-free weighting scheme we introduced in the previous section.

We start by noting that equation (4.17) suggests two feasible weighting schemes:

$$\begin{aligned}(\omega_{vt}^*)^{-1} &= 2 + \frac{n^\alpha Var(\overline{QV}_t - \overline{QV}_{t|t-1})}{2\overline{Q}_t} + \frac{\overline{Q}_{t-1} - \overline{Q}_t}{\overline{Q}_t} \\ &= 2 + \frac{n^\alpha Var(U_t)}{2\overline{Q}_t} + \frac{\overline{Q}_{t-1} - \overline{Q}_t}{\overline{Q}_t}\end{aligned} \tag{4.19}$$

where we use the v subscript refers to the unconditional variance of U, and the subscript u relates to the actual value of U_t, yielding:

$$\begin{aligned}(\omega_{ut}^*)^{-1} &= 2 + \frac{n^\alpha (\overline{QV}_t - \overline{QV}_{t|t-1})^2}{2\overline{Q}_t} + \frac{\overline{Q}_{t-1} - \overline{Q}_t}{\overline{Q}_t} \\ &= 2 + \frac{n^\alpha (\overline{U}_t)^2}{2\overline{Q}_t} + \frac{\overline{Q}_{t-1} - \overline{Q}_t}{\overline{Q}_t}\end{aligned} \tag{4.20}$$

When $\alpha = 1$, in the case of quadratic variation discussed so far, we have:

$$(\omega_{vt}^*)^{-1} = 2 + \frac{nVar(\overline{QV}_t - \overline{QV}_{t|t-1})}{2\overline{Q}_t} + \frac{\overline{Q}_{t-1} - \overline{Q}_t}{\overline{Q}_t} \tag{4.21}$$

$$(\omega_{ut}^*)^{-1} = 2 + \frac{n(\overline{QV}_t - \overline{QV}_{t|t-1})^2}{2\overline{Q}_t} + \frac{\overline{Q}_{t-1} - \overline{Q}_t}{\overline{Q}_t} \tag{4.22}$$

where $\overline{QV}_t = \sum_{j=1}^n r_{t,j}^2$ and $\overline{Q}_t = n/3 \sum_{j=1}^n r_{t,j}^4$ (the appropriate estimator for diffusions without jumps). The distinction between (4.22) and (4.21) will be important when we turn to processes with leverage effect, as will be discussed in section 4.3.3.

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 (4.22) and (4.21). To discuss unconditional weighting schemes we drop time subscripts to the weights ω_t^* in equation (4.11) and consider the generic class of estimators:

$$\overline{QV}_t(\omega) = \overline{QV}_t - \omega(\overline{QV}_t - \overline{QV}_{t|t-1}) \tag{4.23}$$

Recall that QV_t - $QV_{t|t-1} = \sqrt{1-\varphi_0^2}U_t$ and that the relevant trade-off is captured by the product $n^\alpha(1-\varphi_0^2)$, 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 $\overline{QV}_{t|t-1} = \overline{QV}_{t-1}$. Moreover, their approach is unconditional and thus overlooks the difference between Q_t and Q_{t-1}. The unit root case yields the weighting scheme ω^{r-th} = .5 (substituting $\varphi_0 = \varphi = 1$ in equation (4.15) with $Q_t = Q_{t-1}$), and the rule-of-thumb estimator:

$$\overline{QV}_t(\omega^{r-th}) = .5\overline{QV}_t + .5\overline{QV}_{t-1} \tag{4.24}$$

This is a first of two unconditional weighting schemes. Unlike Andreou and Ghysels (2002), Meddahi recognized the trade-off captured by the product $n^\alpha(1-\varphi_0^2)$, and constructed a *model-based* weighting scheme, denoted by $(1-\beta^*)$ and which is characterized as $1-\beta^* = [2+2\lambda]^{-1}$ with:

$$\lambda = n^\alpha[1-\varphi_0]\frac{Var[QV_t]}{2E(Q_t)} \simeq \frac{\gamma^2}{4}\frac{Var[QV_t]}{E(Q_t)} \tag{4.25}$$

It should be noted that Meddahi used the unconditional variance of the estimation error of quadratic variation, that is using our notation $E(Q)/n^\alpha$. 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.[4] To obtain a feasible scheme, we will use unconditional *sample* means of QV_t and Q_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 suggest the second unconditional scheme:

$$
\begin{aligned}
(\omega^{unc})^{-1} &= 2 + \frac{nVar(QV_t - QV_{t|t-1})}{2E(Q_t)} && (4.26)\\
&= 2 + \gamma^2 \frac{Var[QV]}{2E(Q)} && (4.27)\\
&= 2 + \gamma^2 \frac{Var[U_t]}{2E(Q)}
\end{aligned}
$$

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:

$$
[1-\beta^*]^{-1} \simeq 2 + \gamma^2 \frac{Var[QV]}{2E(Q)} \tag{4.28}
$$

with our optimal estimator (4.16), slightly rewritten as:

$$
[\omega^*]^{-1} \simeq (2 + \gamma^2 \frac{Var(U_t)}{2Q_t}) + HQ_t \tag{4.29}
$$

where $HQ_t = (Q_{t-1} - Q_t)/Q_t$ which we referred to as a heteroskedasticity correction. An approach intermediate between (4.29) and (4.28), that is unconditional up to heteroskedasticity correction yields:

$$
\begin{aligned}
(\omega_t^{hc})^{-1} &= 2 + \frac{\overline{Q}_{t-1} - \overline{Q}_t}{\overline{Q}_t} + \gamma^2 \frac{Var[QV_t]}{E(Q_t)} && (4.30)\\
&= 2 + \overline{HQ}_t + \gamma^2 \frac{Var[QV_t]}{2E(Q_t)}
\end{aligned}
$$

The above weighting schemes provides a progression towards the optimal (conditional)

[4] 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 model in our analysis can be any, possibly misspecified, model.

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}. The latter is conditional, yet not fully optimal since $Var[QV]$ is still deflated by the unconditional moment of quarticity. To summarize, we have five possible weights:

$$\omega^{r-th} = 1/2 \tag{4.31}$$

$$\omega^{unc} = [2 + \gamma^2 \frac{Var[QV_t]}{2E(Q_t)}]^{-1} \tag{4.32}$$

$$\omega_t^{hc} = [2 + \overline{HQ}_t + \gamma^2 \frac{Var[QV_t]}{2E(Q_t)}]^{-1} \tag{4.33}$$

$$\omega_{vt}^{*} = [2 + \overline{HQ}_t + \gamma^2 \frac{Var[QV_t]}{2Q_t}]^{-1} \tag{4.34}$$

$$\omega_{ut}^{*} = [2 + \overline{HQ}_t + \gamma^2 \frac{\overline{U}_t^2}{2Q_t}]^{-1} \tag{4.35}$$

Moreover, 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).[5]

- 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 $\overline{Q}_t > \overline{Q}_{t-1}$ that is a larger estimation error on current integrated volatility estimation than on the past one, we may be lead to choose a weight larger than (1/2) for past realized volatility. Typically , taking the term HQ_t into account should do better in the same way WLS are more accurate than OLS in case of conditional heteroskedasticity.

- Besides the heteroskedasticity correction HQ_t we also observe that $Var[QV]/E(Q)$ is replaced by $Var[QV]/Q_t$ in optimal weighting schemes.

4.3 Adding Jumps, Microstructure noise and Leverage

So far we cover a fairly general framework, but restricted our attention to quadratic variation of a diffusion excluding jumps, microstructure noise and leverage. Each of these topics will be discussed here.

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

4.3.1 Bi-Power Variation and Quarticity

In equation (4.1) we allowed for the presence of jumps. In order to separate the jump and continuous sample path components of QV_t Barndorff-Nielsen and Shephard (2004b) and Barndorff-Nielsen and Shephard (2004a) introduce the concept of bi-power variation defined as:

$$\overline{BPV}_t^n(k) = \mu_1^{-2} \sum_{j=k+1}^{n} \left| r_{t,j}^n \right| \left| r_{t,j-k}^n \right|, \tag{4.36}$$

where $\mu_a = E \left| Z \right|^a$ and $Z \sim N(0,1)$, $a > 0$. Henceforth we will, without loss of generality , specialize our discussion the case $k = 1$, and therefore drop it to simplify notation. Barndorff-Nielsen and Shephard (2004b) establish the sampling behavior of $\overline{BPV}_t^n$ as $n \to \infty$, and show that under suitable regularity conditions:

$$\lim_{n\to\infty} \overline{BPV}_t^n(k) = \sigma_t^{[2]}. \tag{4.37}$$

Therefore, in the presence of jumps, $\overline{BPV}_t^n$ converges to the continuous path component of QV_t and is not affected by jumps. The sampling error of the bi-power variation is

$$\frac{n^{\alpha/2}\left(\overline{BPV}_t - \int_{t-1}^{t} \sigma^2(s)\, ds\right)}{\sqrt{\nu_{bb} Q_t}} \sim N(0,1) \tag{4.38}$$

where $\nu_{bb} = (\pi/4)^2 + \pi$ - $5 \approx 0.6090$. Based on these results, Barndorff-Nielsen and Shephard (2004a) and Barndorff-Nielsen and Shephard (2004b) introduce a framework to test for jumps based on the fact that QV consistently estimates the quadratic variation, while $\overline{BPV}$ consistently estimates the integrated variance, even in the presence of jumps. Thus, the difference between the $\overline{QV}$ and the $\overline{BPV}$ is sum of squared jumps (in the limit). Once we have identified the jump component, we can subtract it from the realized variance and we will have the continuous part of the process.

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

$$\overline{BPV}_t(\omega_t^*) = \overline{BPV}_t - \omega_t^*(\overline{BPV}_t - \overline{BPV}_{t|t-1}) \tag{4.39}$$

will be an optimal improvement of $\overline{BPV}_t$ when ω_t^* is again defined according to the following control variable formula (3.12) 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.

For BPV we can summarize the weighting schemes as follows:

$$(\omega_{vt}^{*})^{-1} = 2 + \frac{nVar[\overline{BPV}_t - \overline{BPV}_{t|t-1}]}{\nu \overline{Q}_t} + \frac{\overline{Q}_{t-1} - \overline{Q}_t}{\overline{Q}_t} \tag{4.40}$$

$$(\omega_{ut}^{*})^{-1} = 2 + \frac{n(\overline{BPV}_t - \overline{BPV}_{t|t-1})^2}{\nu \overline{Q}_t} + \frac{\overline{Q}_{t-1} - \overline{Q}_t}{\overline{Q}_t} \tag{4.41}$$

$$(\omega^{unc})^{-1} = = 2 + \frac{nVar[\overline{BPV}_t - \overline{BPV}_{t|t-1}]}{\nu E(Q_t)} \tag{4.42}$$

$$(\omega_{t}^{hc})^{-1} = 2 + \frac{nVar[\overline{BPV}_t - \overline{BPV}_{t|t-1}]}{\nu E(Q_t)} + \frac{\overline{Q}_{t-1} - \overline{Q}_t}{\overline{Q}_t} \tag{4.43}$$

where $\nu = 0.609$, $\overline{BPV}_t = (\pi/2) \sum_{j=2}^{n} |r_{t,j} r_{t,j-1}|$ and $\overline{Q}_t = n(\pi/2)^2 \sum_{j=4}^{n} |r_{t,j} r_{t,j-1} r_{t,j-2} r_{t,j-3}|$. The case of bi-power variation can be generalized to measures involving more general functions, as in Barndorff-Nielsen, Graversen, Jacod, and Shephard (2006). Provided such measure feature persistence we can apply the above analysis in a more general context. One particular case of interest is power variation, typically more persistent than quadratic variation or related measures, as discussed in detail in Forsberg and Ghysels (2006).

4.3.2 Microstructure noise: More general estimators of volatility

In the case if microstructure noise, instead of observing $p(t)$ from (4.1) directly, the price is observed with additive error. This situation has been extensively studied in recent literature. In this case, good estimators $\overline{QV}_t$ have in common that there is still a convergence of the form (4.6), but with different values of α and different definitions of Q_t. In the case of the two scales realized volatility (TSRV) (Zhang, Mykland, and Aït-Sahalia (2005)) $\alpha = 1/6$, and in the case of the multi-scale estimator (MSRV) (Zhang (2006)), or the kernel estimators of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008), $\alpha = 1/4$. The latter rate is efficient since it also occurs in the parametric case (see Gloter and Jacod (2000)). The analysis in the earlier sections for the no-jump case goes through with these modifications.

In the case of the TSRV, $2Q_t$ is replaced by

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

where ν^2 is the variance of the noise, and c is a smoothing parameter. For the more complicated case of the MSRV and the kernel estimators, we refer to the publications cited. We also refer to these publications for a further review of the literature, which includes, in particular, Bandi and Russell (2006) and Hansen and Lunde (2006).

In the case where there are both jumps and microstructure, there are two different targets that can be considered, either the full quadratic variation QV_t, or only its continuous part. (as in the preceding Section 4.3.1).

For estimation of the full quadratic variation, the estimators from the continuous case remain consistent, and retain the same rate of convergence as before. The asymptotic variance $2Q_t$ needs to be modified. The results in this paper for the no-jump case therefore remain valid.

For estimation of the continuous part of QV_t only, there is no fully developed theory. The paper by Fan and Wang (2006) argues for the existence of an $n^{-1/4}$-consistent estimator in the presence of both jumps and noise, but does not display an asymptotic variance. The work by Huang and Tauchen (2006) provides a complete theory, but under the assumption that the microstructure noise is Gaussian.

This paper does not consider the case of infinitely many jumps. There is by now some theory by Aït-Sahalia and Jacod (2004), Aït-Sahalia and Jacod (2006), Woerner (2004), and Woerner (2006) for the situation where there is no microstructure noise.

Irregular observations can be handled using the concept of quadratic variation of time (Mykland and Zhang (2006)).

4.3.3 The case with leverage effect

We here consider the question of how to build a theory in the case where there is leverage effect. In its broadest formulation, what this means is that there is dependence between σ_t and the jumps sizes on the one hand, and the driving Brownian motion and Poisson process on the other. In other words, the analysis cannot be done conditionally on σ_t and the jumps sizes. Equations such as (3.12), (4.10) and (4.12), with their implicit conditioning, are therefore no longer meaningful.

In the no-leverage case, 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, there is no general way of doing the conditioning for a fixed sample size. However, in Appendix D we show 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. Thus stable convergence saves the no-leverage result for the general setting.

It should be noted that for other functionals than estimators of integrated volatility, this phenomenon may no longer hold. The approach, however, can be used in many other setting, see, in particular, the results on estimation of the leverage effect in the next Section 5, where we do the relevant calculations explicitly.

5 Estimating the Leverage Effect

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 into its own when trying to estimate less well posed quantities. By way of example, we consider how to estimate the leverage effect. The concept of leverage effect is used to cover several concepts, but we here take it to mean the covariance $L_t = \langle p, \sigma^2 \rangle_t - \langle p, \sigma^2 \rangle_{t-1}$ in the model (4.1). (For simplicity, consider the no-jump case, $\lambda = 0$, which can be undone as in Section C.2). Specifically, if $dW_t = \rho_t dW_1(t) + \sqrt{1-\rho_t^2} dW_2(t)$, and the system is given by

$$dp\,(t) = \mu\,(t)\,dt + \sigma\,(t)\,dW\,(t)$$
$$d\sigma^2 = \nu(t)dt + \gamma(t)dW_1(t),$$

(where all of $\rho(t)$, $\nu(t)$, and $\gamma(t)$ can be random processes, we obtain that

$$L_t = \int_{t-1}^{t} \sigma(t)\gamma(t)\rho(t)dt.$$

Leverage effect is also used to mean the correlation $\rho(t)$ between p and σ^2, or, in general, the dependence between σ^2 and the Brownian motion W.

An estimate of leverage effect is given by

$$\hat{L}_t = 2\sum_i (\hat{\sigma}^2_{\tau_{n,i+1}} - \hat{\sigma}^2_{\tau_{n,i}})(p_{\tau_{n,i+1}} - p_{\tau_{n,i}}), \tag{5.1}$$

where

$$\hat{\sigma}^2_{\tau_{n,i}} = \frac{T}{n(M-1)} \sum_{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}]} (\Delta p_{t_{n,j+1}} - \overline{\Delta p}_{\tau_{n,i}})^2 \tag{5.2}$$

and $\overline{\Delta p}_{\tau_{n,i}} = (p_{\tau_{n,i}} - p_{\tau_{n,i-1}})/M$, where $t_{n,j} = t - 1 + j/n$ and $\tau_{n,i} = t - 1 + iM/n$. This estimate is given in Section 4.3 of Mykland and Zhang (2009), where it is shown that the estimate (for fixed M, as $n \to \infty$) is asymptotically unbiased, but not consistent. To be precise,

$$\hat{L}_t - L_t \to v_{M,t}^{1/2} Z_t, \tag{5.3}$$

in law, where the Z_t are (independent) standard normal, and

$$v_{M,t} = \frac{16}{M-1}\left(\frac{2M}{2M+3}\right)^2 \int_{t-1}^{t} \sigma_u^6 du. \tag{5.4}$$

It is conjectured that if $M \to \infty$ as $n \to \infty$, specifically $M = O(n^{1/2})$, then the estimator will be consistent with an $O_p(n^{-1/4})$ rate of convergence, but the conjecture also suggests that, for practical data sizes, M has to be so large relative to n that little is gained relative to (5.3) by considering the consistent version.

We are now in a situation, therefore, where high frequency data is not quite as good at providing information about the underlying quantity to be estimated. If we take the fixed M estimator as our point of departure, we do not even need to make triangular array type assumptions like (3.18) for our procedure to make sense asymptotically. If we let $\varphi_L \equiv Cov(\hat{L}_t, \hat{L}_{t-1})/Var(\hat{L}_t)$ (in analogy with (3.8) and (3.9)), we can let the optimal linear predictors of $\hat{L}_t$ given $\hat{L}_{t-1}$, be written as $\hat{L}_{t|t-1}$= $(1-\varphi_L)L + \varphi_L \hat{L}_{t-1}$, in analogy with (3.8). Again, here L is the unconditional unbiased time series mean of L_t.

A combined linear estimator of L_t is thus $\hat{L}_t(\omega_t) = \hat{L}_t$- $\omega_t(\hat{L}_t$ - $\hat{L}_{t|t-1})$, where we note that, as $n \to \infty$

$$\hat{L}_{t|t-1} \to (1-\varphi_L)L + \varphi_L L_{t-1} + \phi v_{M,t-1}^{1/2} Z_{t-1},$$

in law, and so, again in law,

$$\hat{L}_t(\omega_t) - L_t \to (1-\omega_t) v_{M,t}^{1/2} Z_t + \omega_t \left[(1-\varphi_L)L + \varphi_L L_{t-1} - L_t + \varphi_L v_{M,t-1}^{1/2} Z_{t-1} \right].$$

The asymptotic MSE (conditional on the data) is therefore:

$$MSE_c = (1-\omega_t)^2 v_{M,t} + \omega_t^2 \left[((1-\varphi_L)L + \varphi_L L_{t-1} - L_t)^2 + \varphi_L^2 v_{M,t-1} \right].$$

The (infeasible) optimal value ω_t^* is thus

$$\omega_t^* = \frac{v_{M,t}}{(\varphi_L(L_{t-1} - L) - (L_t - L))^2 + \varphi_L^2 v_{M,t-1} + v_{M,t}}. \tag{5.5}$$

In this case, therefore, there is no need for φ_L to go to 1 as $n \to \infty$.

6 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}. While the simulations using 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, see however Ghysels, Mykland, Renault, and Wang (2009).

6.1 Simulation design

We consider 1,000 replications of samples each consisting of 500 and 1,000 'days' with in-sample (intra-daily sampling) sizes $n = 288$, 144 and 24. These correspond to the use of five-minute, ten-minute and hourly 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 first class of models we simulate are based on Andersen, Bollerslev, and Meddahi (2005) and consist of:

$$\begin{aligned} d\log\ S_t &= \mu dt + \sigma_t dW_t \\ &= \sigma_t[\rho_1 dW_{1t} - \rho_2 dW_{2t} + \sqrt{1-\rho_1^2-\rho_2^2}dW_{3t}] \end{aligned} \tag{6.1}$$

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

$$d\log\ S_t = \sigma_t dW_{3t} \tag{6.2}$$

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

$$d\sigma_t^2 = .035(.636 - \sigma_t^2)dt + .144\sigma_t^2 dW_{1t} \tag{6.3}$$

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

$$\begin{aligned} d\sigma_{1t}^2 &= .5708(.3257 - \sigma_{1t}^2)dt + .2286\sigma_{1t}^2 dW_{1t} \\ d\sigma_{2t}^2 &= .0757(.1786 - \sigma_{2t}^2)dt + .1096\sigma_{2t}^2 dW_{2t} \end{aligned} \tag{6.4}$$

All of the above models satisfy the regularity conditions of the Jacod (1994) and Barndorff-Nielsen and Shephard (2002) 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 do not report the 1-hour sampling frequency, instead focusing on the most commonly used sampling frequencies of 5 and 10 minute high frequency data.[6]

[6] All unreported results are available upon request from the authors.

For the purpose of option pricing we also consider a second class of models - namely Heston-type models. Here, we consider 1000 simulations of 2000 days samples following a first 1000 days burn-in pre-sample. The values of parameters in Heston model are taken from Chernov and Ghysels (2000), namely, $r = 5.814\%$, $\kappa = 0.6901$, $\theta = 0.0096$, $\sigma = 0.0615$ and $\rho = -0.0183$. They are divided by 252 to convert to a daily scale.

Finally, all our simulations are based on AR(1) prediction schemes for the purpose of constructing the weights. Hence, all our simulations consider looking at the previous day's QV only.

6.2 Simulation results

Tables 1 and 2 report the MSE comparisons summary statistics for the weights for QV, with $(\omega_{vt}^{*})^{-1}$ appearing in equation (4.21), $(\omega_{ut}^{*})^{-1}$ appearing in equation (4.22), and $(\omega_{t}^{hc})^{-1}$ from equation (4.30), and $(\omega^{unc})^{-1}$ from equation (4.26), where $QV_t = \sum_{j=1}^{n} r_{t,j}^2$ and $Q_t = n/3 \sum_{j=1}^{n} r_{t,j}^4$, without jumps. Note that we only consider the case without jumps, since we know the DGPs do not feature jumps. Table 1 pertains to the GARCH diffusion, whereas Table 2 covers the two-factor diffusion. In the tables we report the mean improvement across all simulations (Mean), the Monte Carlo simulation variance (Var) and for the weights we also report the Monte Carlo mean of the sample variance of the weights (meanvar).

We observe that overall $(\omega_{vt}^{*})^{-1}$ appears to be the best weighting scheme in terms of MSE improvement, as it reduces by more than 20 to 25 % the mean squared error for GARCH diffusions and the two factor model. The latter figures pertain to the 10 minute sampling interval where the differences between that alternative weighting schemes is the most pronounced. It is interesting to note that $(\omega_{ut}^{*})^{-1}$ appears to perform poorly, not only with respect to $(\omega_{vt}^{*})^{-1}$ but also with respect to $(\omega_{t}^{hc})^{-1}$, where the latter is the natural progression between $(\omega_{vt}^{*})^{-1}$ and $(\omega^{unc})^{-1}$. The rule-of-thumb scheme ω^{r-th} has the interesting property that it is both the best and worst, depending on the DGP. For the GARCH diffusion it outperforms $(\omega_{ut}^{*})^{-1}$, and $(\omega_{vt}^{*})^{-1}$, while for the two-factor model it is worse than the unfiltered RV series. We also decomposed the MSE into squared bias and variance. The results are easy to summarize, as in all cases the bias is trivially small. Hence, the MSE comparisons are essentially efficiency comparisons.

It is worth recalling that $(\omega_{vt}^{*})^{-1}$ is more closely related to what would be the Kalman filter weights. The bottom panels of both tables display summary statistics for the weights for the sample size $T = 500$ (the large sample size is omitted as it yields similar results). The average weight for $(\omega_{ut}^*)^{-1}$ is as high as .25, whereas for $(\omega_{vt}^*)^{-1}$ it is only .16 for the GARCH diffusion - with similar results for the two factor model. Note also that the weights of $(\omega_t^{hc})^{-1}$ and $(\omega^{unc})^{-1}$ are somewhat similar. These results show there are clear differences between $(\omega_{ut}^*)^{-1}$ and $(\omega_{vt}^*)^{-1}$ and it seems that for filtering the latter performs better.

The conclusions regarding the ranking of weights change when we turn our attention to forecasting gains discussed in subsection 2.3. We consider two experiments. The first is a linear forecasting exercise:

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

where the regressor X_t is either one of the following: IV_t, QV_t and corrected QV_t using the optimal weighting $(\omega_{vt}^*)^{-1}$ and $(\omega_{ut}^*)^{-1}$ schemes with one-day lag of information. Obviously the infeasible benchmark is the regression involving IV as regressor. Hence, we compare how close the feasible raw QV_t and corrected measure perform in comparison. We do this for 5 and 10 minute sampling schemes.

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

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

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

The simulation evidence is reported in Table 3 and focuses exclusively on the second experiment described above - as the first one did not yield many differences between the various estimation schemes. There is now a clear difference between the ranking in Tables 1 and 2 and that in Table 3. In the case of filtering, we find that $(\omega_{ut}^*)^{-1}$ is slightly better than $(\omega_{vt}^*)^{-1}$. Hence, the conditional weighting scheme with realized U_t, i.e. $(\omega_{ut}^*)^{-1}$ outperforms $(\omega_{vt}^*)^{-1}$, and therefore as far as forecasting a nonlinear function goes, the ranking is reversed.

Although the differences are not as significant, there is a clear pattern of dominance.

7 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 - see Ghysels, Mykland, Renault, and Wang (2009).

Appendix

A Proofs of Propositions 3.1 and 3.2

We first check that Assumption 3.2 ensures weak stationarity of the process $(\hat{\sigma}_t^2)$ and endows it with second order characteristics conformable to those listed after Assumption 3.2. From the conditional moments given after Assumption 3.1, we deduce:

$$\begin{aligned} E(\hat{\sigma}_t^2) &= E[E_c(\hat{\sigma}_t^2)] = E(\sigma_t^2) \\ Var(\hat{\sigma}_t^2) &= Var[E_c(\hat{\sigma}_t^2)] + E[Var_c(\hat{\sigma}_t^2)] = Var(\sigma_t^2) + \frac{2}{n}E[\sigma_t^{[4]}] \end{aligned}$$

Similarly, since:

$$\hat{\sigma}_t^2 = \sum_{j=1}^{n} [r_{t,j}^n]^2$$

the process $(\hat{\sigma}_t^2)$ is serially independent given the volatility path and for all $h > 0$:

$$Cov[\hat{\sigma}_t^2, \hat{\sigma}_{t-h}^2] = Cov[E_c(\hat{\sigma}_t^2), E_c(\hat{\sigma}_{t-h}^2)] = Cov[\sigma_t^2, \sigma_{t-h}^2]$$

We can now prove the result of Proposition 3.1. It should first be noted that:

$$\hat{\sigma}_t^2 - \hat{\sigma}_{t|t-1}^2 = \hat{\sigma}_t^2 - \varphi\hat{\sigma}_{t-1}^2 - (1-\varphi)\sigma^2 \qquad \text{(A.1)}$$

and therefore:

$$\begin{aligned} Var_c[\hat{\sigma}_t^2 - \hat{\sigma}_{t|t-1}^2] &= Var_c[\hat{\sigma}_t^2 - \varphi\hat{\sigma}_{t-1}^2] \\ &= \frac{2\sigma_t^{[4]}}{n} + \frac{2\varphi^2\sigma_{t-1}^{[4]}}{n} \end{aligned}$$

From equation (A.1) we also obtain that:

$$E_c[\hat{\sigma}_t^2 - \hat{\sigma}_{t|t-1}^2] = \sigma_t^2 - \varphi\sigma_{t-1}^2 - (1-\varphi)\sigma^2$$

Finally, using the same equation we have:

$$\begin{aligned} Cov_c[\hat{\sigma}_t^2, \hat{\sigma}_t^2 - \hat{\sigma}_{t|t-1}^2] &= Cov_c[\hat{\sigma}_t^2, \hat{\sigma}_t^2 - \varphi\hat{\sigma}_{t-1}^2] \\ &= \frac{2\sigma_t^{[4]}}{n} \end{aligned}$$

Using equation (3.12) and collecting all the above results we obtain:

$$\omega_t^* = \frac{2\sigma_t^{[4]}/n}{2/n[\sigma_t^{[4]} + \varphi^2\sigma_{t-1}^{[4]}] + [\sigma_t^2 - \varphi\sigma_{t-1}^2 - (1-\varphi)\sigma^2]^2}$$

and hence equation (3.13).

Next we turn to Proposition 3.2. In order to prove the Proposition we use the above formulas on second order characteristics of the weakly stationary process $(\hat{\sigma}_t^2)$ to deduce:

$$\varphi = \frac{Cov[\hat{\sigma}_t^2\ , \hat{\sigma}_{t-1}^2\ ]}{Var(\hat{\sigma}_t^2)} \tag{A.2}$$

$$= \frac{Cov(\sigma_t^2, \sigma_{t-1}^2)}{Var\sigma_t^2 + (2/n)(E[\sigma_t^{[4]}])} \tag{A.3}$$

which can also be written as:

$$\varphi = \frac{\varphi_0 Var\sigma_t^2}{Var\sigma_t^2 + (2/n)(E[\sigma_t^{[4]}])}$$

dividing denominator and numerator by $Var\sigma_t^2$ yields equation (3.16).

Finally, let us discuss the validity of the approximation (3.20) and its consequence in the block-local-to-unity context (3.21). By a Rieman integration argument, we have almost surely:

$$Lim_{n=\infty}\frac{1}{n}\sum_{j=1}^{n}\sigma_{t,j}^2 = \int_{t-1}^{t}\sigma^2(u)du \tag{A.4}$$

where $[\sigma^2(u)]_{t-1\leq u\leq t}$ is a continuous time process such that:

$$\sigma^2(t-1+\frac{j}{n}) = \sigma_{t,j}^2, 1 \leq j \leq n$$

Assuming it is a stationary diffusion process, the process $[\sigma^2(u)]_{t-1\leq u\leq t}$ is conformable with the model (3.19) (with $\rho = \exp(-k\Delta t)$) if and only if it has a linear drift with a mean-reversion parameter k:

$$d\sigma^2(t) = k(\sigma^2 - \sigma^2(t))dt + \gamma(t)dW(t)$$

where $W(t)$ is a Wiener process. Note that an Ornstein-Uhlenbeck-like process as put forward by Barndorff-Nielsen and Shephard (2001) would also work. 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. We can in particular 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:

$$\frac{[1 - \exp(-k)]^2}{2[\exp(-k) - 1 + k]}$$

Since we can assume without loss of generality the needed necessary conditions for validity of a dominated convergence theorem, we then deduce from the convergence (A.4) that, for n large, the correlation coefficient between $\sigma_t^2 = \frac{1}{n}\sum_{j=1}^{n} \sigma_{t,j}^2$ and $\sigma_{t-1}^2 = \frac{1}{n}\sum_{j=1}^{n} \sigma_{t-1,j}^2$ is given by the same formula:

$$\varphi_0 \sim \frac{[1 - \exp(-k)]^2}{2[\exp(-k) - 1 + k]}$$

Using $k \sim \frac{d}{n}$, a Taylor expansion gives:

$$\varphi_0 \sim [1 - \frac{k}{2}]^2 \sim 1 - k \sim 1 - \frac{d}{n}$$

B A Comparison of Two Estimators

The unbiased estimator defined in equation (3.24) can be rewritten as,

$$\begin{aligned} \hat{\sigma}_t^{[4]} &= \frac{m}{n} \sum_{t=1}^{m/n} \frac{\hat{\sigma}_{t,[i]}^4}{1 + 2/m} \\ &= \frac{n}{m+2} \sum_{i=1}^{n/m} [\sum_{j=m(i-1)+1}^{j=mi} r_{t,j}^2]^2 \end{aligned} \tag{B.1}$$

The above estimator can be compared with the naive estimator appearing in (3.25). To do so we need to derive the conditional variance of $\hat{\sigma}_t^{[4]}$. Note that we can rewrite the estimator (3.24) as:

$$\begin{aligned} \hat{\sigma}_t^{[4]} &= \frac{1}{(m+2)n} \sum_i \sigma_{t,[i]}^4 [\sum_j (\frac{r_{t,j}}{\sigma_{t,[i]}/\sqrt{n}})^2]^2 \\ &= \frac{1}{(m+2)n} \sum_i \sigma_{t,[i]}^4 [\chi_i^2(m)]^2 \end{aligned}$$

Since $E\left[[\chi_i^2(m)]^{p/2}\right] = 2^{p/2}\Gamma((p+m)/2)/\Gamma(m/2)$, therefore $E[\chi_i^2(m)]^4 = 2^4\ \Gamma(4+m/2)/\Gamma(m/2)$ $= 2^4\ (3+m/2)\ (2+m/2)\ (1+m/2)\ m/2$. Consequently, $E[\chi_i^2(m)]^4 = (m+6)(m+4)(m+2)m$. Along similar lines, one has $E[\chi_i^2(m)]^2 = m(m+2)$. Therefore,

$$\begin{aligned} Var[\chi_i^2(m)^2] &= (m+6)(m+4)(m+2)m - m^2(m+2)^2 \\ &= 8m(m+2)(m+3) \end{aligned}$$

The above results yield:

$$\begin{aligned} Var[\hat{\sigma}_t^{[4]}] &= \frac{1}{n^2(m+2)^2} \sum_i \sigma_{t,[i]}^8 8m(m+2)(m+3) \\ &= \frac{8m(m+3)}{n^2(m+2)} \sum_i \sigma_{t,[i]}^8 \end{aligned}$$

We now turn our attention to the naive estimator written as in equation (3.25):

$$\begin{aligned} \tilde{\sigma}_t^{[4]} &= \frac{1}{3n} \sum_i \sigma_{t,[i]}^4 \sum_j (\frac{r_{t,j}}{\sigma_{t,[i]}})^4 \\ &= \frac{1}{3n} \sum_i \sigma_{t,[i]}^4 \sum_j (\chi^2(1))^2 \end{aligned}$$

Therefore,

$$\begin{aligned} Var(\tilde{\sigma}_t^{[4]}) &= \frac{1}{9n^2} \sum_i \sigma_{t,[i]}^8 Var((\chi^2(1))^2) \times m \\ &= \frac{32m}{3n^2} \sum_i \sigma_{t,[i]}^8 \end{aligned}$$

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 $\hat{\sigma}_t^4$. First, note that

$$\hat{\sigma}_t^4 = \frac{1}{n(m+2)} \sum_{i=1}^{m/n} \sigma_{t,[i]}^4 \varepsilon_i^2 \tag{B.2}$$

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

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

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

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

and hence,

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

C From Discrete to Continuous Time

To make the connection to Section 3, assume first that there are no jumps ($\lambda \equiv 0$), that $\mu_t \equiv 0$, and also that the σ_t process is independent of W_s, so that there is no leverage effect. In this case, one can carry out inference conditional on the σ_t process, and still retain the structure

$$dp\,(t) = \sigma\,(t)\,dW\,(t) \tag{C.1}$$

where W is a Brownian motion. p_t is now a conditionally Gaussian process, and Assumption 3.1 is satisfied, by making the identification

$$\sigma_{t,j}^2 = \int_{t-(j-1)/n}^{t-j/n} \sigma_s^2 ds \tag{C.2}$$

Furthermore, one can even make the stronger Assumption 3.3, without substantial loss of generality.

The precise rationale for this is as follows. *Define*

$$\sigma_{t,[i]}^2 = \frac{1}{m} \sum_{j=(i-1)m+1}^{im} \sigma_{t,j}^2. \tag{C.3}$$

Now compare two probability distributions: (1) P^* is given by Assumption 3.1, and (2) P_n given by Assumption 3.3, with the link provided by (C.3). As is shown in Theorem 1 of Mykland (2006), P^* and P_n are measure-theoretically equivalent, and they are also contiguous in the strong sense that dP_n/P^* converges in law under P^* to a random variable with mean 1. The same is true for dP^*/dP_n relative to P_n.

The consequence of this is that estimators that are consistent under Assumptions 3.1 or 3.3 remain consistent for the general process C.1. Furthermore, all rates of convergence are preserved, as well as asymptotic variances. To be more precise, for an estimator $\hat{\theta}_n$, if $n^{1/2}(\hat{\theta}_n - \theta) \to N(0, a^2)$ under P_n, then $n^{1/2}(\hat{\theta}_n - \theta) \to N(b, a^2)$ under P^* and under (C.1). The only modification is therefore a possible bias b, which has to be characterized in each individual case.[7]

In summary, the development in Section 3 remains valid for the model C.1 so long as all variances and MSEs are interpreted asymptotically.

C.1 Reinstating μ

Having shown that the development in Section 3 covers the simplified model (C.1), we now argue that it also covers the more general case (4.1). To see this, consider first the case where $\lambda \equiv 0$. Call P the probability distribution of the process $p(t)$ under (4.1), while P^* is the probability distribution of the process under (C.1). In this case, it follows from Girsanov's Theorem that P and P^* are, subject to weak regularity conditions, measure-theoretically equivalent. Once again, this means that consistency and orders or convergence are preserved from P^* to P. Also, as in the case of μ and λ equal zero, the asymptotic normal distribution of $n^{1/2}(\hat{\theta}_n - \theta)$ is preserved, with the same variance, but possibly with a bias that has to be found in each special case. In the case of the estimators of volatility (quadratic variation, bi-power) and of quarticity that are considered in this paper, this bias is zero. The general theory is discussed in Section 2.2 of Mykland and Zhang (2009).

[7]In the case of the estimators of volatility (quadratic variation, bi-power) and of quarticity that are considered in this paper, it is easy to see that $b = 0$. For general estimators, we refer to Mykland (2006), and Mykland and Zhang (2009).

C.2 Reinstating λ

The conceptually simplest approach is to remove these jumps before further analysis. Specifically, declare a *prima facie* jump in all intervals $(t-1+(j-1)/n, t-1+j/n]$ with absolute return $|p_{t-1+j/n} - p_{t-1+(j-1)/n}| > \log n/n^{1/2}$. Provisionally remove these intervals from the analysis. Then carry out the approximation described as in the case of μ and λ equal zero on the remaining intervals.

The procedure will detect all intervals $(t-1+(j-1)/n, t-1+j/n]$, with probability tending to one (exponentially fast) as $n \to \infty$. If one simply removes the detected intervals from the analysis, it is easy to see that our asymptotic results go through unchanged. The intervals where jumps have been detected must be handled separately.

To give a concrete example of how the approach works, consider the bi-power sum. Let I_n be the intervals with a detected jump. Write

$$\sum_{j=1}^{n} \left|r_{t,j}^n\right| \left|r_{t,j-1}^n\right| = \sum_{j,j-1 \notin I_n} \left|r_{t,j}^n\right| \left|r_{t,j-1}^n\right| + \sum_{j \text{ or } j-1 \in I_n} \left|r_{t,j}^n\right| \left|r_{t,j-1}^n\right| \tag{C.4}$$

The first term on the left hand side of (C.4) can now be handled as in Section 3; the second term is handled separately and directly. Since there are only finitely many such terms, this is straightforward.

The procedure is like the one described in Mancini (2001) and Lee and Mykland (2008). See also Aït-Sahalia and Jacod (2007). Here, however, it is used only for purposes of analysis, and not for actual estimation.

D The case of leverage

The point of departure is that the convergence (4.6) remains valid even under leverage effect, as shown in Section 5 of Jacod and Protter (1998) and Proposition 1 of Mykland and Zhang (2006) . Specifically, suppose that the underlying filtration is generated by a p-dimensional local martingale $(\chi^{(1)}, \ldots, \chi^{(p)})$. It is then the case that

$$n^{\alpha/2}(\overline{QV}_t - QV_t) \xrightarrow{d} Z_t\sqrt{2Q_t}, \tag{D.1}$$

where Z_t is standard normal, and the convergence is joint with $(\chi^{(1)}, \ldots, \chi^{(p)})$ (where this is a constant sequence). Z_t is independent of $(\chi^{(1)}, \ldots, \chi^{(p)})$ (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, Q_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 (D.1) in hand, one can now condition the asymptotic distribution on the data (*i.e.*, $(\chi^{(1)}, \ldots, \chi^{(p)})$), and obtain that $Z_t\sqrt{2Q_t}$ is (conditionally) normal with mean zero and variance $2Q_t$.

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 $\varphi(n)$ (cfr. equation (3.17) 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:

$$QV_t = \varphi_0(n)QV_{t-1} + \sqrt{1-\varphi_0(n)^2}U_t(n) + (1-\varphi_0(n))E(QV_t), \tag{D.2}$$

One supposes that in the limit as $n \to \infty$, QV_{t-1} and U_t are uncorrelated. A similar, feasible, equation is then written as

$$\overline{QV}_t(n) = \varphi(n)\overline{QV}_{t-1}(n) + \sqrt{1-\varphi(n)^2}\overline{U}_t(n) + (1-\varphi(n))E(QV_t), \tag{D.3}$$

Specifically, in analogy with (4.14),

$$n^\alpha(1-\varphi_0(n)^2) = \gamma_0^2 \text{ and } n^\alpha(1-\varphi(n)^2) = \gamma^2. \tag{D.4}$$

Note that

$$1-\varphi(n) = n^{-\alpha}\frac{1}{2}\gamma^2(1+o_p(1)) = O_p(n^{-\alpha}). \tag{D.5}$$

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 (D.2) from (D.3), and then multiplies by $n^{\alpha/2}$, (D.1) yields

$$\begin{aligned} n^{\alpha/2}(\overline{QV}_t(n) - QV_t(n)) &= \varphi(n)n^{\alpha/2}(\overline{QV}_{t-1}(n) - QV_{t-1}(n)) + n^{\alpha/2}(\varphi(n)-\varphi_0(n))QV_{t-1}(n) \\ &+ n^{\alpha/2}\sqrt{1-\varphi(n)^2}\overline{U}_t(n) - n^{\alpha/2}\sqrt{1-\varphi_0(n)^2}U_t(n) - n^{\alpha/2}(\varphi(n)-\varphi_0(n))E(QV_t) \\ &\to \sqrt{2Q_{t-1}}Z_{t-1} + \gamma\overline{U}_t - \gamma_0 U_t, \end{aligned} \tag{D.6}$$

whence

$$\gamma\overline{U}_t + \sqrt{2Q_{t-1}}Z_{t-1} = \gamma_0 U_t + \sqrt{2Q_t}Z_t, \tag{D.7}$$

hence U_t is not even asymptotically observable.

Under this setup, from (D.5),

$$n^{\alpha/2}(\overline{QV}_t - \varphi(n)\overline{QV}_{t-1}) \to \gamma\overline{U}_t. \tag{D.8}$$

Consider the best linear forecast of $\overline{QV}_t$ using (only) $\overline{QV}_{t-1}$:

$$\overline{QV}_{t|t-1} = \varphi(n)\overline{QV}_{t-1} + (1-\varphi(n))E(QV)$$

so that from (D.5)

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

The final estimate is now $\overline{QV}_t(\omega_t) = \overline{QV}_t - \omega_t(\overline{QV}_t - \overline{QV}_{t|t-1})$, hence

$$\begin{aligned} n^{\alpha/2}(\overline{QV}_t(\omega_t) - QV_t) &= n^{\alpha/2}(\overline{QV}_t - QV_t) - \omega_t n^{\alpha/2}(\overline{QV}_t - \overline{QV}_{t|t-1}) \\ &\to \sqrt{2Q_t}Z_t - \omega_t\gamma\overline{U}_t \\ &= (1-\omega_t)\sqrt{2Q_t}Z_t - \omega_t\left[-\sqrt{2Q_{t-1}}Z_{t-1} + \gamma_0 U_t\right] \end{aligned} \tag{D.9}$$

by (D.7). Hence, the asymptotic MSE (conditional on the data) is

$$MSE_c = (1-\omega_t)^2 2Q_t + \omega_t^2\left[2Q_{t-1} + \gamma_0^2 U_t^2\right] \tag{D.10}$$

The stable convergence (D.1) remains valid even in this triangular array setup by invoking Proposition 3 of Mykland and Zhang (2006).

Under assumption (D.2), one can therefore do the same calculations as before, but on asymptotic quantities. The result (4.15) then remains valid: the asymptotic mean squared error (conditional on the data) of the overall estimate $QV_t(\omega_t)$ is minimized by

$$\omega_t^* = \frac{Q_t}{Q_t + Q_{t-1} + \gamma_0^2 U_t^2/2}. \tag{D.11}$$

The further development is the same as in the no-leverage case.

References

Aït-Sahalia, Y., and J. Jacod (2004): "Fisher's Information for Discretely Sampled Lévy Processes," Discussion paper, Princeton University and Université de Paris 6.

——— (2006): "Volatility Estimators for Discretely Sampled Lévy Processes," *Annals of Statistics, forthcoming.*

——— (2007): "Volatility Estimators for Discretely Sampled Lévy Processes," *Annals of Statistics*, 35, 335–392.

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

Andersen, T., T. Bollerslev, and F. Diebold (2009): "Parametric and Nonparametric Volatility Measurement," in *Handbook of Financial Econometrics, (forthcoming).*, ed. by Y. Aït-Sahalia, and L. P. Hansen.

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

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

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.

Bandi, F., J. Russell, and C. Yang (2008): "Realized volatility forecasting and option pricing," *Journal of Econometrics*, 147, 34–46.

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

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

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

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.

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

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

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

——— (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., R. Engle, and D. Nelson (1994): "ARCH Models," in *Handbook of Econometrics*. Elsevier Science, Amsterdam.

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

Chernov, M., and E. Ghysels (2000): "A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation," *Journal of Financial Economics*, 56, 407–458.

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

Fan, J., and Y. Wang (2006): "Multi-scale Jump and Volatility Analysis for High-Frequency Financial Data," Discussion paper, University of Connecticut.

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

Foster, D. P., and D. B. Nelson (1996): "Continuous record asymptotics for rolling sample variance estimators," *Econometrica*, 64(1), 139–174.

Ghysels, E., A. Harvey, and E. Renault (1996): "Stochastic Volatility," in *Handbook of statistics, Vol. 14, Statistical Methods in Finance*. North Holland.

Ghysels, E., P. A. Mykland, E. Renault, and F. Wang (2009): "Testing for jumps with pre-weighting," University of Chicago and University of North Carolina.

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

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 (2008): "Edgeworth Corrections for Realized Volatility," *Econometric Reviews*, 27, 139–162.

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

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

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

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

Jacod, J. (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.

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

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., AND L. ZHANG (2006): "ANOVA for Diffusions and Itô Processes," *The Annals of Statistics*, 34, 1931–1963.

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

MYKLAND, P. A. (2006): "A Gaussian Calculus for Inference from High Frequency Data," Annals of Finance (forthcoming).

PHILLIPS, P., H. MOON, AND Z. XIAO (2001): "How to estimate autoregressive roots near unity," *Econometric Theory*, 17, 29–69.

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

SHEPHARD, N. (2004): *Stochastic Volatility - Selected Readings*. Oxford University Press.

STOCK, J. (1991): "Confidence intervals of the largest autoregressive root in US macroe conomic time series," *Journal of Monetary Economics*, 28, 435–459.

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

WOERNER, J. H. (2004): "Purely discontinuous Lévy processes and power variation: Inference for integrated volatility and the scale parameter," Discussion paper, University of Göttingen.

——— (2006): "Power and multipower variation: Inference for high frequency data," in *Stochastic Finance*, ed. by A. Shiryaev, M. do Rosário Grosshino, P. Oliviera, and M. Esquivel, pp. 264 – 276. Springer Verlag, Berlin.

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

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

ZHANG, L., P. 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 1: MSE improvements and weights GARCH Diffusion Model

MSE improvements with AR(1) Prediction

	MSE raw	Impr: ω^*_{vt}		Impr: ω^*_{ut}		Impr: ω^{r-th}		Impr: ω^{unc}		Impr: ω^{hc}_t	
	Mean	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var
Sample size: 500 days											
1 Min	0.0007	0.95	0.0081	0.95	0.0006	2.93	0.1488	0.92	0.0003	0.93	0.0003
5 Min	0.0038	0.81	0.0052	0.87	0.0004	0.99	0.0134	0.86	0.0005	0.84	0.0004
10 Min	0.0077	0.74	0.0033	0.83	0.0005	0.73	0.0055	0.79	0.0005	0.77	0.0003
Sample size: 1000 days											
1 Min	0.0008	0.99	0.0141	0.97	0.0003	2.93	0.1030	0.94	0.0003	0.94	0.0001
5 Min	0.0038	0.81	0.0043	0.89	0.0006	0.97	0.0081	0.85	0.0004	0.84	0.0004
10 Min	0.0078	0.75	0.0029	0.87	0.0003	0.75	0.0035	0.80	0.0002	0.79	0.0003

Weights with AR(1) prediction

	ω^{unc}		ω^*_v			ω^*_u			ω^{hc}_t		
	Mean	Var	Mean	Var	meanvar	Mean	Var	meanvar	Mean	Var	meanvar
Sample size=500 days											
1 Min	0.076	0.00004	0.068	0.00005	0.00311	0.184	0.00051	0.02502	0.095	0.00004	0.00001
5 Min	0.174	0.00013	0.141	0.00027	0.00831	0.217	0.00059	0.02394	0.192	0.00013	0.00012
10 Min	0.209	0.00015	0.167	0.00038	0.01042	0.251	0.00070	0.02378	0.214	0.00015	0.00039

Table 2: MSE improvements and weights Two Factor Diffusion Model

MSE improvements with AR(1) Prediction

	MSE raw	Impr: ω^*_{vt}		Impr: ω^*_{ut}		Impr: ω^{r-th}		Impr: ω^{unc}		Impr: ω^{hc}_t	
	Mean	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var
					Sample size: 500 days						
1 Min	0.0003	0.96	0.0005	0.99	0.0002	6.25	0.5912	0.97	0.0001	0.97	0.0001
5 Min	0.0020	0.87	0.0010	0.93	0.0003	1.65	0.0277	0.89	0.0002	0.89	0.0002
10 Min	0.0039	0.79	0.0011	0.89	0.0003	1.07	0.0091	0.84	0.0002	0.83	0.0003
					Sample size: 1000 days						
1 Min	0.0003	0.96	0.0003	0.99	0.0001	6.22	0.3001	0.97	0.0001	0.97	0.0001
5 Min	0.0019	0.86	0.0006	0.93	0.0002	1.64	0.0166	0.89	0.0001	0.89	0.0001
10 Min	0.0039	0.78	0.0005	0.89	0.0002	1.07	0.0056	0.83	0.0001	0.82	0.0001

Weights with AR(1) prediction

	ω^{unc}		ω^*_v			ω^*_u			ω^{hc}_t		
	Mean	Var	Mean	Var	meanvar	Mean	Var	meanvar	Mean	Var	meanvar
					Sample size=500 days						
1 Min	0.041	0.00001	0.071	0.00001	0.00054	0.145	0.00006	0.02325	0.055	0.00001	0.00001
5 Min	0.127	0.00006	0.120	0.00005	0.00359	0.188	0.00088	0.02626	0.166	0.00006	0.00007
10 Min	0.176	0.00007	0.162	0.00007	0.00603	0.224	0.00005	0.02632	0.194	0.00007	0.00032

Table 3: Forecasting Log Black-Scholes Implied Volatilities - Heston Model

We consider:

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

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

X_t		IV	QV Corrected(ω^*_{vt})	QV Corrected(ω^*_{ut})	QV
			R^2 1-day ahead forecast, ATM		
TTM: 22days	5 Min	0.98	0.93	0.94	0.93
	10 Min	0.98	0.88	0.91	0.90
TTM: 44days	5 Min	0.98	0.91	0.94	0.93
	10 Min	0.98	0.86	0.91	0.90
TTM: 66days	5 Min	0.98	0.93	0.94	0.93
	10 Min	0.98	0.92	0.94	0.93
			R^2 5-day ahead forecast, ATM		
TTM: 22days	5 Min	0.91	0.86	0.88	0.87
	10 Min	0.91	0.82	0.85	0.84
TTM: 44days	5 Min	0.91	0.86	0.88	0.87
	10 Min	0.91	0.82	0.85	0.84
TTM: 66days	5 Min	0.91	0.85	0.87	0.87
	10 Min	0.91	0.82	0.85	0.84
			R^2 20-day ahead forecast, ATM		
TTM: 22days	5 Min	0.72	0.68	0.70	0.69
	10 Min	0.72	0.65	0.67	0.66
TTM: 44days	5 Min	0.72	0.68	0.69	0.69
	10 Min	0.72	0.65	0.67	0.67
TTM: 66days	5 Min	0.72	0.68	0.69	0.69
	10 Min	0.72	0.64	0.670	0.67