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

Eric Ghysels
University of North Carolina†

Per Mykland
University of Chicago‡

Eric Renault
University of North Carolina§

First Draft: December 2006

This Draft: November 25, 2008

Preliminary and Incomplete

Abstract

We revisit the widely used in-sample asymptotic analysis developed by Jacod (1994) and Barndorff-Nielsen and Shephard (2002) and 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 of Bobkoski (1983), Phillips (1987), Chan and Wei (1987) among many others.

*The authors thank 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 as well as Nour Meddahi and Neil Shephard for comments.

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

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

§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 process 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 underpinnings of a now standard two-step modelling 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 (2002). 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 modelling, which is the subject of the second step. It is argued

[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).

in this paper that measuring volatility is not necessarily confined to a single (daily) interval and prior observations are useful thanks to persistence in volatility. 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 revisited Andreou and Ghysels (2002) 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 paper is organized as follows....

2 Main motivation

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. 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. 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), Phillips (1987) 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. The link with local-to-unity asymptotics is a key part of our derivations, both allowing us to remain model-free as far as filter weights go, and allowing us to be conditional on the volatility path. It should also be noted that the arguments provided so far do not only apply to volatility measurement, but are also relevant for higher moment measurements, 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 estimators for the cumulants.

Second, we do in fact not necessarily have to rely on local-to-unity asymptotic arguments. For example, existing estimators for the measurement of leverage effects using intra-daily data are not consistent albeit asymptotically unbiased, see e.g. Mykland and Zhang (2007). 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.

3 An Introductory 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. 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:

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

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 $\sigma^2_{t,j}$, $j = 1, \ldots, n$, the return process in equation (3.1) is distributed independently Gaussian:*

$$r^n_{t,j} \sim N(0, \frac{1}{n}\sigma^2_{t,j}) \tag{3.2}$$

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

For every period t the parameter of interest is:

$$\sigma^2_t \equiv Var(\sum_{j=1}^{n}[r^n_{t,j}]) \equiv \frac{1}{n}\sum_{j=1}^{n}\sigma^2_{t,j} \tag{3.3}$$

and consider the following ML estimators for each t :

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

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

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

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

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

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^2_t using prior day information. After characterizing the optimal weighting scheme we discuss estimation issues in subsection 3.2.

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 $\hat{\sigma}_{t-1}^2$. To formalize this, assume that:

Assumption 3.2 *The process* (σ_t^2) *is weakly stationary.*

Assumption 3.2 implies that we can engage in regression analysis, and in particular consider:

$$\varphi \equiv \frac{Cov(\hat{\sigma}_t^2, \hat{\sigma}_{t-1}^2)}{Var(\hat{\sigma}_t^2)} \tag{3.7}$$

$$\varphi_0 \equiv \frac{Cov(\sigma_t^2, \sigma_{t-1}^2)}{Var(\sigma_t^2)} \tag{3.8}$$

where φ and φ_0 are respectively the coefficients of the optimal linear predictors of $\hat{\sigma}_t^2$ given $\hat{\sigma}_{t-1}^2$, and σ_t^2 given σ_{t-1}^2. It is important to note the difference between equations (**??**) and (3.7). In the former case, the process is contaminated by estimation noise, whereas in the latter case we have the time series regression using population quantities.

From the above discussion it is clear that there are potentially two types of asymptotics. On the one hand, the number of observations n per period t can increase to infinity. On the other hand, the number of time periods $t = 1, \ldots, T$ can also become large. The asymptotics we handle in the paper is the former. Regarding the time series process, our setting is standard and relies on the usual assumptions. We will therefore ignore the uncertainty pertaining to time series estimation:

Assumption 3.3 *The parameters* φ *is estimated via sample equivalents of (**??**), denoted* $\hat{\varphi}_T$*, and we assume that* T/n *is infinitely large such that all time series estimation error can be ignored, since* $Var_c(\hat{\sigma}_t^2) = O(1/n)$.

While Assumption 3.3 implies we no longer have to worry about the difference between φ and $\hat{\varphi}_T$, it will still be the case that we need to distinguish φ from φ_0 since the latter results from estimation errors of order $1/n$ separating the population quantity σ_t^2 and its sample counterpart $\hat{\sigma}_t^2$. For the moment we will proceed with estimated quantities $\hat{\sigma}_t^2$, and hence φ. Later we will discuss the relationship between φ and φ_0.

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

$$\hat{\sigma}_{t|t-1}^2 = (1-\varphi)\sigma^2 + \varphi\hat{\sigma}_{t-1}^2 \tag{3.9}$$

where σ^2 is the (unbiased) unconditional (time series) mean of $\hat{\sigma}_t^2$ and of σ_t^2 as well. The goal is to combine $\hat{\sigma}_{t|t-1}^2$ and $\hat{\sigma}_t^2$, to improve the estimation of σ_t^2 using prior day information. In particular, we define a new estimator combining linearly $\hat{\sigma}_{t|t-1}^2$ and $\hat{\sigma}_t^2$:

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

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:

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

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 $(\hat{\sigma}_t^2 - \hat{\sigma}_{t|t-1}^2)$ 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 (4.15). 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, both our Monte Carlo experiments and empirical work 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 empirical 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 unfeasible bias:

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

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

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

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

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 B

Thus since:

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

it is equivalent to show that the feasible bias or the unfeasible one will not necessarily dominate the conditional standard error $2\sigma_t^{[4]}/n + 2\varphi^2\sigma_{t-1}^{[4]}/n$. The logic of our approach starts from the observation that this dominance is not necessary maintained, provided that the persistence parameter φ_0 is 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 volatility process to ensure that the forecast $\hat{\sigma}^2_{t|t-1}$ of σ^2_t - which uses past information - still improves $\hat{\sigma}^2_t$, which uses n intra-daily observations. This approach is reminiscent of local-to-unity asymptotics of Bobkoski (1983), Phillips (1987) and Chan and Wei (1987) among many others.[2] Note however an important difference between the original local-to-unity asymptotics and our use of it. While the former near-to-unit root literature 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:

$$[1 - \varphi_0^2(n)] = O(\varphi_0(n) - \varphi(n))$$

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 of (3.19) 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:

$$(B_t^F)^2 = O([1 - \varphi_0^2(n)]) = O(\varphi_0(n) - \varphi(n)) \tag{3.19}$$

does not dominates the conditional variance $Var_c(\hat{\sigma}^2_t - \hat{\sigma}^2_{t|t-1})$.

3.2 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.

[2]A recent extension to block local-to-unity asymptotics by Phillips, Moon, and Xiao (2001) has some resemblance with our analysis, although we focus here on integrated volatility.

Because of Assumption 3.3 we do not need to worry about the latter, hence our focus will be on the former.

We first turn to the 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 3.4 *Assume that n is a multiple of $m \geq 1$, and for $(i-1)m < j \leq im$, we have:*

$$\sigma_{t,j} = \sigma_{t,[i]} \qquad i = 1,\ldots,n/m$$

Given Assumption 3.4 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:

$$\hat{\sigma}_t^{[4]} = \frac{m}{n}\sum_{i=1}^{n/m}\frac{\hat{\sigma}_{t,[i]}^4}{1+2/m} \tag{3.20}$$

$$= \frac{n}{m+2}\sum_{i=1}^{n/m}\left[\sum_{j=m(i-1)+1}^{mi} r_{t,j}^2\right]^2 \tag{3.21}$$

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

$$\begin{aligned}
\tilde{\sigma}_t^{[4]} &= \frac{n}{3} \sum j = 1^n r_{t,j}^4 \\
&= \frac{n}{3} \sum_i \sigma_{t,[i]}^4 \sum_j (\frac{r_{t,j}}{\sigma_{t,[i]}})^4 \quad (3.22) \\
&\backsim \frac{1}{3n} \sum_i \sigma_{t,[i]}^4 (\chi^2(1))^2 \quad (3.23)
\end{aligned}$$

In Appendix C 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, consider by analogy:

$$\Psi = \frac{Cov(\hat{\sigma}_t^{[4]}, \hat{\sigma}_{t-1}^{[4]})}{Var(\hat{\sigma}_t^{[4]})} \quad (3.24)$$

Since,

$$\begin{aligned}
\hat{\sigma}_{t,[i]}^4 &= \sigma_{t,[i]}^4 \frac{[\chi_i^2(m)]^2}{m^2} \\
&= \sigma_{t,[i]}^4 \varepsilon_i^2
\end{aligned}$$

In Appendix C we also compute the unconditional variance

$$Var \hat{\sigma}_t^4 = Var[\sigma_t^4] + 2 \frac{8m(m+3)}{m^2(m+2)} E[\sigma_{t,[i]}^8] \quad (3.25)$$

Which allows us to write

$$\Psi = \frac{\Psi_0}{1 + 2 \frac{8m(m+3)}{m^2(m+2)} \frac{\sum_{i=1}^{n/m} E[\sigma_{t,[i]}^8]}{Var[\sigma_t^4]}} \quad (3.26)$$

4 General Theory

In the previous section we started with a relatively simple example of a piecewise constant volatility process. We consider in this section the case of general jump diffusion models. The theory in this section is asymptotic in nature in terms of sampling of intra-daily data as well as the properties of the data generating process across days. Regarding the data generating process, we rely on a new local-to-unity asymptotic argument which consists of assuming that the persistence across days is sufficiently high to achieve a asymptotic trade-off with the intra-daily sampling frequency going to infinity. In a first subsection 4.1 we show the theory developed in section 3 applies to the general case of jump diffusions without leverage. Subsection 4.3 covers the estimation of quadratic variation without jumps followed by a subsection 4.4 distinguishing conditional weighting schemes discussed so far and unconditional schemes suggested in some of the prior literature. Subsection 4.5 covers bi-power variation and quarticity, that is statistics measuring quadratic variation in the presence of jumps and high order moments. More general projections in the construction of weighting scheme appear in subsection 4.6. We then turn to the topic of forecasting with improved measurement in subsection 4.7. A final subsection 4.8 covers the issue of microstructure noise.

4.1 From Discrete to Continuous Time

The example in the previous section is surprisingly comprehensive. We here explain why, and then develop a general theory.

We start with a continuous time stochastic volatility jump-diffusion model for asset returns, namely:

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

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

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

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

4.1.1 The Reduced Form Problem: no μ, no λ

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 ones inference conditionally of the σ_t process, and still retain the structure

$$dp(t) = \sigma(t)\, dW(t) \tag{4.3}$$

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

$$\sigma_{t,j}^2 = \int_{t-(j-1)/n}^{t-j/n} \sigma_s^2 ds \tag{4.4}$$

Furthermore, one can even make the stronger Assumption 3.4, without substantial loss of generality. The precise rationale for this is as follows. *Define*

$$\sigma_{t,[i]}^2 = \frac{1}{m} \sum_{j=(i-1)m+1}^{im} \sigma_{t,j}^2. \tag{4.5}$$

Now compare two probability distributions: P^* is given by Assumption 3.1, and P_n is given by Assumption 3.4, with the link provided by (4.5). 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 3.4 remain consistent for the general process 4.3. Furthermore, all rates of convergence are preserved, as well as asymptotic variances. To be precise, for an estimator $\hat{\theta}_n$, if $n^{1/2}(\hat{\theta}_n - \theta) \rightarrow N(0, a^2)$ under P_n, then $n^{1/2}(\hat{\theta}_n - \theta) \rightarrow N(b, a^2)$ under P^* and under 4.3. The only modification is therefore a possible bias b, which has to be found in each individual case.

In the case of the estimators of volatility (quadratic variation, bipower) 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 (2007).

In summary, the development in Section 3 remains valid for the model 4.3 so long as all variances and MSEs are interpreted asymptotically. The precise results are given in Section 4.2-4.3.2.

4.1.2 Reinstating μ

Having shown that the development in Section 3 covers the simplified model (4.3), 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 (4.3). 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, just as in Section 4.1.1, 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, bipower) 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 (2007).

4.1.3 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 $|p_{t-1+j/n} - p_{t-1+(j-1)/n}| > \log n/n^{1/2}$. [Verify] Provisionally remove these intervals from the analysis. Then carry out the approximation described in Section 4.1.1 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 bipower sum. Let I_n

be the intervals with a detected jump. Write

$$\sum_{j=1}^{n} \left| r_{t,j}^{n} \right| \left| r_{t,j-1}^{n} \right| = \sum_{j,j-1 \notin I_n} \left| r_{t,j}^{n} \right| \left| r_{t,j-1}^{n} \right| + \sum_{j \text{ or } j-1 \in I_n} \left| r_{t,j}^{n} \right| \left| r_{t,j-1}^{n} \right| \tag{4.6}$$

The first term on the left hand side of (4.6) 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 (2006). See also Aït-Sahalia and Jacod (2007). Here, however, it is used only for purposes of analysis, and not for actual estimation.

4.2 The Continuous Time Case

We now discuss the implications of the previous discussion for the estimation of quadratic variation (integrated volatility) for model (4.1): $dp\,(t) = \mu\,(t)\,dt + \sigma\,(t)\,dW\,(t) + \kappa\,(t)\,dq\,(t)$.

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. The intra-daily return is then denoted $r_{t,j}^{n} = p_{t-j/n} - p_{t-(j-1)/n}$ where $1/n$ is the (intra-daily) sampling frequency. For example, when dealing with typical stock market data we will use $n = 78$ corresponding to a five-minute sampling frequency. It is possible to consistently estimate QV_t in (4.2) by summing squared intra-daily returns, yielding the so called realized variance, namely:

$$\overline{QV}_{t}^{n} = \sum_{j=1}^{n} \left(r_{t,j}^{n} \right)^{2}. \tag{4.7}$$

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.

$$\lim_{n \to \infty} \overline{QV}_{t}^{n} \to^{p} QV_{t}. \tag{4.8}$$

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

$$\frac{\sqrt{n}(\overline{QV}_t - QV_t)}{\sqrt{2Q_t}} \xrightarrow{d} N(0,1) \tag{4.9}$$

where $Q_t = \int_{t-1}^{t} \sigma(s)^4 ds$ is called the quarticity.

We note that in the case of no leverage effect, the result 4.9 follows from the simple example in Section 3 in view of the discussion in Section 4.1. The case with leverage is discussed in 4.9.

We will consider first the case of quadratic variation without jumps, discussed in a first subsection 4.3. In subsection 4.3.1 we study infeasible estimation, whereas in subsection 6.1 we study a specific example of a one-factor stochastic volatility process. Finally, subsection **??** extends our analysis to processes with jumps and processes contaminated by microstructure noise.

4.3 The case of Quadratic Variation without Jumps

For simplicity we will focus first on the case without jumps. 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 lagged day of observations but for the moment, we simplify the exposition to what amounts to an order one Markov process. 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. For purpose of later extensions it will be useful to parameterize the rates of convergence with α. Setting $\alpha = 1$, we have:

$$\begin{aligned} \frac{n^{\alpha/2}(\overline{QV}_{t-1} - QV_{t-1})}{\sqrt{2Q_{t-1}}} &\xrightarrow{d} N(0,1) \\ \frac{n^{\alpha/2}(\overline{QV}_t - QV_t)}{\sqrt{2Q_t}} &\xrightarrow{d} N(0,1) \end{aligned} \tag{4.10}$$

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

[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 $(\sigma_s^2)_{s\in[t-1,t]}$.

coefficient:

$$\varphi = \frac{Cov(QV_t, QV_{t-1})}{Var(QV_{t-1})} \tag{4.11}$$

and the unconditional expectation $E(QV_t) = E(QV_{t-1}) = E(QV)$. Note that equation (4.11) does *not* imply that our analysis is confined to AR(1) models. Instead, equation (4.11) only reflects the fact that we condition predictions on a single lag QV_{t-1}. There may potentially be 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 models with a single lag. Higher order models are a straightforward extension that will be discussed later.

The theory presented in the following will mirror the development in Section 3, but be valid when volatility is not constant across a day.

Consider the best linear forecast of QV_t using (only) QV_{t-1} :

$$QV_{t|t-1} = \varphi QV_{t-1} + (1-\varphi)E(QV)$$

and to compute its realized counterpart:

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

Of course, 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: $\overline{QV_\tau}$, $\tau = t-1, ..., t-T+1$. This will introduce two additional issues: (i) estimation error of φ and $E(QV)$ that would have been obtained if we had observed QV_τ, $\tau = t-1, ..., t-T+1$, (ii) additional estimation error due to the fact that we only observe $\overline{QV_\tau}$, $\tau = t-1, ..., t-T+1$. While, as far as integrated volatility is concerned, the former error will be made negligible by assuming (T/n^α) goes to infinity, the latter may not be negligible and is the subject of the paper.

4.3.1 Infeasible estimation

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:

$$\overline{QV}_t(\omega_t) = (1-\omega_t)\overline{QV}_t + \omega_t \overline{QV}_{t|t-1} \tag{4.12}$$

Intuitively, the more persistent the volatility process, the more $QV_{t|t-1}$ is informative about 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:

$$E[\overline{QV}_t(\omega_t) - QV_t]^2 = E\{\overline{QV}_t - QV_t - \omega_t(\overline{QV}_t - \overline{QV}_{t|t-1})\}^2 \tag{4.13}$$

Then, it can be shown that:

$$\overline{QV}_t(\omega_t^*) = \overline{QV}_t - \omega_t^*(\overline{QV}_t - \overline{QV}_{t|t-1}) \tag{4.14}$$

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

$$\omega_t^* = \frac{Cov[\overline{QV}_t, \overline{QV}_t - \overline{QV}_{t|t-1}]}{Var(\overline{QV}_t - \overline{QV}_{t|t-1}) + [E(\overline{QV}_t - \overline{QV}_{t|t-1})]^2} \tag{4.15}$$

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.

To do this, we want to apply Lemma 3.1, by computing moments given the volatility path. More precisely, we write:

$$\overline{QV}_t - \overline{QV}_{t|t-1} = (\overline{QV_t} - QV_t) + (QV_t - QV_{t|t-1}) - \varphi(\overline{QV}_{t-1} - QV_{t-1})$$

Then, given the volatility path, we have:

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

Therefore, ω_t^* defined by (4.15) can be rewritten as:

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

Note that for the case of realized volatility, equation (4.15) is a corollary to Proposition 3.1, in view of the discussion is Section 4.1.

In order to estimate volatility on day t, we thus give 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 $(QV_t - QV_{t|t-1})^2$. However, for a given non-zero forecast error, the optimal weight ω_t^* goes to 0 when n goes to infinity. The reason for that is very clear: 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 $(QV_t - QV_{t|t-1})^2$. To correctly assess the latter, it is worth noting that:

$$QV_t - QV_{t|t-1} = \sqrt{1-\varphi^2} u \tag{4.16}$$

where, by a simple argument of variance decomposition, the variable u has a zero unconditional mean and an unconditional variance equal to $Var(QV_t) = Var(QV_{t-1}) = Var(V)$. Therefore:

$$\omega_t^* = \frac{Q_t}{Q_t + \varphi^2 Q_{t-1} + n^\alpha (1-\varphi^2) u^2/2} \tag{4.17}$$

The relevant trade-off is then clearly captured by the product $n^\alpha(1-\varphi^2)$. The trade-off is sensible because daily integrated volatility will be sufficiently persistent (φ 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 $\varphi(n)$. More precisely, and in analogy with the assumption (3.17), let us assume that for some given number γ we have:

$$n^\alpha (1 - \varphi(n)^2) = \gamma^2 \tag{4.18}$$

Hence, we have a drifting Data Generating Process ($\varphi(n)$ increasing with n) to capture the idea that, the larger n is, the larger volatility persistence $\varphi(n)$ 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. Then the optimal weight is:

$$\omega_t^* = \frac{Q_t}{Q_t + (1 - \frac{\gamma^2}{n^\alpha})Q_{t-1} + \gamma^2 u^2/2} + o(\frac{1}{n^\alpha}) = \frac{Q_t}{Q_t + Q_{t-1} + \gamma^2 u^2/2} + O(\frac{1}{n^\alpha}) \qquad (4.19)$$

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.

Note that 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. While the level of the optimal weights depend on n, it should be noted that the temporal dependence of the weights also depends on n, because the temporal dependence of Q_t depends $\varphi(n)$. 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.

4.3.2 Feasible estimation

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.

Besides the estimation of quarticity we face another problem. Consider equation (4.16) and the resulting weighting scheme (4.17). Combining this with (4.18) yielded (4.19). Since

$Var[QV_t] = Var[u]$, we can rewrite equation (4.19) in terms of inverse of optimal weights as:

$$[\omega_t^*]^{-1} = 2 + \frac{Q_{t-1} - Q_t}{Q_t} + \gamma^2 \frac{Var[QV_t]}{Q_t} \tag{4.20}$$

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 $(1-\varphi^2)^{1/2}u$ in equation (4.19). First, we should note that we don't observe $QV_t - QV_{t|t-1} = (1-\varphi^2)^{1/2}u$ 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 $(1-\varphi^2)^{1/2}u$. Therefore, as a proxy we may consider the following feasible estimator:

$$[\omega_t^{f*}]^{-1} = 2 + \frac{Q_{t-1} - Q_t}{Q_t} + \gamma^2 \frac{(\overline{QV}_t - \overline{QV}_{t|t-1})^2}{(1-\varphi^2)Q_t} \tag{4.21}$$

$$= 2 + \frac{Q_{t-1} - Q_t}{Q_t} + \frac{(\overline{QV}_t - \overline{QV}_{t|t-1})^2}{Q_t/n^\alpha} \tag{4.22}$$

The optimal weighting scheme ω_t^{f*} will be used extensively in our simulation study as well as in the comparison with unconditional weighting schemes discussed in the next section.

4.4 A comparison with unconditional rules

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.

To discuss unconditional weighting schemes we drop time subscripts to the weights ω_t^* in equation (4.32) and consider the generic class of estimators:

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

We noted in section 4 that QV_t - $QV_{t|t-1} = \sqrt{1-\varphi^2}u$ and that the relevant trade-off is captured by the product $n^\alpha(1-\varphi^2)$, which led us to use 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 $\varphi = 1$, which amounts de facto to a unit root case and therefore $\overline{QV}_{t|t-1} = QV_{t-1}$. The unit root case yields the weighting scheme $\omega^{r-th} = .5$ (substituting $\varphi = 1$ in equation (4.17)), and the rule-of-thumb estimator:

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

Meddahi, instead did recognize the trade-off, and constructed a *model-based* weighting scheme, denoted by $(1-\beta^*)$ and which is characterized as: $1-\beta^* = [2+2\lambda]^{-1}$ and where:

$$\lambda = n^\alpha[1-\varphi]\frac{Var[QV]}{E(Q)} \simeq \frac{\gamma^2}{2}\frac{Var[QV]}{E(Q)} \tag{4.25}$$

It should be noted that Meddahi used $Var[u_t(h)]$, which is the unconditional variance of the estimation error of quadratic variation, that is using our notation $E(Q)/n^\alpha$. Moreover, he assumes an explicit data generating process to compute the weights, hence a model is needed to be specified (and estimated) to compute the weights.[4] The above derivations allows us to compare:

$$[1-\beta^*]^{-1} \simeq 2 + \gamma^2\frac{Var[QV]}{E(Q)} \tag{4.26}$$

with our optimal estimator (4.32), slightly rewritten as:

$$[\omega^*]^{-1} \simeq (2 + \gamma^2\frac{u^2}{Q_t}) + HQ_t \tag{4.27}$$

where $HQ_t = (Q_{t-1} - Q_t)/Q_t$, which we will refer to later as a heteroskedasticity correction. From the above analysis we can make several observations:

- The formula in equation (4.26) of Meddahi gives a weight to past realized volatility smaller than the rule-of-thumb weight of (1/2).[5]
- Equation (4.26) does not take into account the conditional heteroskedasticity that is

[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.

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

due to the (asymptotic) estimation error of realized volatility. For instance, when $Q_t > Q_{t-1}$ 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 $Var[QV]/Q_t$.

To appraise the differences between the weighting schemes we will consider various schemes based on equation (4.23) with:

$$(\omega^{unc})^{-1} = 2 + \gamma^2 \frac{Var[QV]}{E(Q)} \tag{4.28}$$

$$(\omega^{unc-hc})^{-1} = 2 + \frac{Q_{t-1} - Q_t}{Q_t} + \gamma^2 \frac{Var[QV]}{E(Q)}$$

and where in practice the term $(\gamma^2 Var[QV])/E(Q)$ will be again 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 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 feasible version of ω^{unc} will be denoted ω^{f-unc}. Likewise, we use the sample mean to compute $\omega^{f-unc-hc}$.

The weighting schemes in (4.28) represent a natural progression towards the optimal (conditional) weighting scheme ω^* derived in the previous section 4. Starting with the rule-of-thumb scheme ω^{r-th}, we progress to ω^{f-unc} where unconditional moments are used, followed by the heteroskedasticity correction embedded in $\omega^{f-unc-hc}$. The latter is already conditional, yet not fully optimal since $Var[QV]$ is still deflated by the unconditional moment of quarticity. Finally, in the simulations we will compare these three weighting schemes with our optimal *feasible* weighting scheme (4.21).

4.5 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 bipower variation defined as:

$$\overline{BPV}_t^n(k) = \mu_1^{-2} \sum_{j=k+1}^{n} \left| r_{t,j}^n \right| \left| r_{t,j-k}^n \right|, \tag{4.29}$$

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:

$$\lim_{n\to\infty} \overline{BPV}_t^n(k) = \sigma_t^{[2]}. \tag{4.30}$$

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

$$\frac{n^{\alpha/2}\left(\overline{BPV}_t - \int_{t-1}^{t} \sigma^2(s)\, ds\right)}{\sqrt{\nu_{bb} Q_t}} \sim N(0,1) \tag{4.31}$$

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:

$$\overline{BPV}_t(\omega_t^*) = \overline{BPV}_t - \omega_t^*(\overline{BPV}_t - \overline{BPV}_{t|t-1}) \tag{4.32}$$

will be an optimal improvement of $\overline{BPV}_t$ when ω_t^* is again defined according to the following control variable formula (4.15) 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.

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).

4.6 More general projections

So far we confined our analysis to projections on one lag. We start with generalizing the models we considered so far. We consider higher order projections. It might be useful to think of ADF representation to accommodate the local-to-unity asymptotics, following Stock (1991):

$$\Delta QV_t = \varphi_0 + \varphi QV_{t-1} + \sum_{i=1}^{p-1} \varphi_i \Delta QV_{t-i} + \varepsilon_t \tag{4.33}$$

where φ is the sum of the autoregressive coefficients φ_i, $i = 1$,..., p. Following Stock (1991) we can apply local-to-unity asymptotics to the sum of AR coefficients, i.e. make $\varphi(n)$ a function of n, as we did in the AR(1) case.

Incomplete

4.7 Forecasting

So far we focused exclusively 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 and the purpose of this subsection is to analyze the impact of our improvement measurements on forecasting performance.

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, for general purposes we denote the process to forecast as 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 observations: $Z_{t+1} = (1-\omega)X_{t+1} + \omega Y^*_{t+1}$, where Y^*_{t+1} is an unbiased linear predictor of X_{t+1}. Note that this refers to the unconditional schemes discussed earlier in subsection 4.4. Since Y^*_{t+1} is unbiased linear predictor of X_{t+1} :

$$X_{t+1} = Y^*_{t+1} + v^*_{t+1}, E(v^*_{t+1}) = 0, Cov[v^*_{t+1}, Y^*_{t+1}] = 0$$

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

$$X_{t+1} = Y^X_{t+1} + v^X_{t+1}, E(v^X_{t+1}) = 0, Cov[v^X_{t+1}, Y^X_{t+1}] = 0$$

It is natural to assume in addition that:

$$Cov[v^X_{t+1}, Y^*_{t+1}] = 0$$

hence, the predictor Y^*_{t+1} does not allow us to improve the preferred predictor Y^X_{t+1}. Consider now a modified forecaster due to the improvement measurement, and let us denote it by:

$$Y^Z_{t+1} = (1-\omega)Y^X_{t+1} + \omega Y^*_{t+1}$$

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

$$Var(Y^X_{t+1} - Y_{t+1}) - Var(Y^Z_{t+1} - Y_{t+1}) = \omega^2(Var(v^X_{t+1}) - Var(v^*_{t+1})) \tag{4.34}$$

This result has following implications: 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).

Some caveats are in order about our discussion so far. Namely, our forecasting exercise only involves fixed weights and linear forecasting rules. As it was shown in the previous subsection 4.4 and further documented later simulation we know that conditional optimal weights are far superior to the unconditional ones. Therefore, one should expect the forecasting gains to be more important with conditional weights. In fact, the conditional weighting schemes result de facto in nonlinear prediction formulas due to the time variation in the weights.

In many volatility forecasting application the object of interest is a nonlinear function of future volatility. The most prominent example is option pricing, where the future path of volatility until time to maturity determines the current option pricing through a conditional expectation of a nonlinear payoff function. A simplified example is the model of **?** where the price of a European call option of time to maturity h and moneyness k equals:

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

where $BS(\sigma^2, k, h)$ 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. Note that the simplifying assumption regarding leverage and risk pricing should alter these conclusions.

4.8 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 (4.10), 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 (2006), $\alpha = 1/4$. 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 futher 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 4.5).

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)).

4.9 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 (4.11), (4.13) and (4.15), with their implicit conditioning, are therefore no longer meaningful.

The point of departure is that the convergence (4.10) 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 $(\chi^{(1)}, ..., \chi^{(p)})$. It is them the case that

$$n^{\alpha/2}(\overline{QV}_t - QV_t) \xrightarrow{d} Z_t\sqrt{2Q_t}, \tag{4.35}$$

where Z_t is standard normal, and the convergence is joint with $(\chi^{(1)}, ..., \chi^{(p)})$ (where this is a constant sequence). Z_t is independent of $(\chi^{(1)}, ..., \chi^{(p)})$ (the latter also occurs in the limit, since the sequence is constant as a function of n). This is known as stable convergence, cf. 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 $t = 0, ..., T$. In this case, $Z_0, ..., Z_T$ are iid.

With the convergence (4.35) in hand, one can now condition the asymptotic distribution on the data (*i.e.*, $(\chi^{(1)}, ..., \chi^{(p)})$), 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. One writes (3.17) as before, and assumes that

$$QV_t = \varphi_0 QV_{t-1} + \sqrt{1-\varphi_0^2}U_t + (1-\varphi_0)E(QV_t), \tag{4.36}$$

where all the above quantities implicitly depend on n. Specifically, in analogy with (4.18),

$$n^{\alpha}(1-\varphi_0(n)^2) = \gamma_0^2 \tag{4.37}$$

under which

$$n^{\alpha/2}(QV_t - \varphi_0(n)QV_{t-1}) \to \gamma_0 U_t. \tag{4.38}$$

Consider the best linear forecast of QV_t using (only) QV_{t-1} :

$$QV_{t|t-1} = \varphi_0 QV_{t-1} + (1-\varphi_0)E(QV)$$

so that

$$n^{\alpha/2}(QV_t - QV_{t|t-1}) \to \gamma_0 U_t.$$

and to compute its realized counterpart:

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

If we take $\varphi - \varphi_0 = O_p(n^{-\alpha})$, then

$$n^{\alpha/2}(\overline{QV}_{t|t-1} - QV_{t|t-1}) \to Z_{t-1}\sqrt{2Q_{t-1}}$$

so that

$$\begin{aligned} n^{\alpha/2}(\overline{QV}_t - \overline{QV}_{t|t-1}) &= n^{\alpha/2}(\overline{QV}_t - QV_t) - n^{\alpha/2}(\overline{QV}_{t|t-1} - QV_{t|t-1}) + n^{\alpha/2}(QV_t - QV_{t|t-1}) \\ &\to \sqrt{2Q_t}Z_t - \sqrt{2Q_{t-1}}Z_{t-1} + \gamma_0 U_t. \end{aligned}$$

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

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

Hence, the asymptotic MSE (conditional on the data) is

$$MSE_c = (1-\omega_t)^2 2Q_t + \omega_t^2\left[2Q_{t-1} + \gamma_0^2 U_t^2\right] \tag{4.39}$$

One supposes that in the limit as $n \to \infty$, QV_{t-1} and U_t are uncorrelated. The stable convergence (4.35) remains valid even in this triangular array setup by invoking Proposition 3 of Mykland and Zhang (2006).

Under assumption (4.36), one can therefore do the same calculations as before, but on asymptotic quantities. The result (4.19) then remains valid: the asymptotic mean squared error (conditional on the data) of the overall estimate $QV_t(\omega_t)$ is minimized by

$$\omega_t^* = \frac{Q_t}{Q_t + Q_{t-1} + \gamma_0^2 U_t^2/2}. \tag{4.40}$$

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

To summarize the difference between this procedure and the earlier one: 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, 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 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 here 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 4.1.3). Specifically, if $dW_t = \rho_t dW_1(t) + \sqrt{1-\rho_t^2} dW_2(t)$, and the system is given by

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

(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

$$\hat{L}_t = \frac{2M}{2M+3} \sum_i (\hat{\sigma}^2_{\tau_{n,i+1}} - \hat{\sigma}^2_{\tau_{n,i}})(p_{\tau_{n,i+1}} - p_{\tau_{n,i}}), \tag{5.1}$$

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 (2007), where it is shown that the estimate (for fixed M, as $n \to \infty$) is asymptotically unbiased, but not consistent. To be precise,

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

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

$$v_{M,t} = \frac{4}{M-1} \left(\frac{2M}{2M+3} \right)^2 \int_{t-1}^{t} \sigma^6_u du. \tag{5.4}$$

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 (3.7)), 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 (3.9). 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 $n \to \infty$

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

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

6.1 The case of One-factor Stochastic Volatility processes

The analysis so far used a linear prediction model for QV. We did *not* assume this prediction model $\overline{QV}_{t|t-1}$ is the true data generating process, only a particular prediction model. So far we only considered Markov models of order one. In the next subsection we will expand the setting to more general prediction models. Before we do, we consider the case where the first order Markov prediction coincides with the best linear predictor, i.e. the true data generating process is such that the linear AR(1) is the best. This setting allows us to further illustrate the efficiency gains. We consider two example, (1) a one-factor SV model linear in the drift as studied by Meddahi and Renault (2004) and (2) a class of non-Gaussian Ornstein-Uhlenbeck (henceforth OU) processes highlighted in Barndorff-Nielsen and Shephard (2001). In both cases we also exclude again the presence of jumps at this point.

Since we are dealing with a order one autoregressive process with persistence depending on n, it is not surprising that the analysis in this section is reminiscent of the local-to-unity asymptotics commonly used to better approximate the finite sample behavior of parameter estimates in AR(1) models with root near the unit circle where neither the Dickey-Fuller asymptotics nor the standard normal 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. The local-to-unity approach was first proposed by Bobkoski (1983), and subsequently studied by Phillips (1987) and Chan and Wei (1987) among many others.[6]

INSERT FIRST CASE HERE

A non-Gaussian Ornstein-Uhlenbeck process is defined as:

$$d\sigma(t)^2 = -\lambda\sigma(t)^2 dt + dz(\lambda t) \tag{6.1}$$

where $z(t)$ is a Lévy process with non-negative increments and hence allows for jumps in volatility. The specification $dz(\lambda t)$ allows one to keep the marginal distribution of volatility invariant to the choice of λ, provided it is non-negative. This process yields an autocorrelation function $acf(\sigma(t)^2, s) \equiv cor(\sigma(t)^2, \sigma(t+s)^2)$ equal to $acf(\sigma(t)^2, s) = \exp(-\lambda|s|)$. Using results from Barndorff-Nielsen and Shephard one obtains:

$$acf(QV_t, s) = \frac{(1-e^{-\lambda})^2 e^{-\lambda(|s|-1)}}{2(e^{-\lambda} - 1 + \lambda)} \tag{6.2}$$

Hence, the first order autocorrelation denoted φ equals $(1-e^{-\lambda})^2/2(e^{-\lambda} - 1 + \lambda)$.

In the previous subsection it was recognized that $\varphi(n)$ depends on the sampling frequency. Using equation (4.18) and the second order Taylor expansion of the exponential function we have:

$$\lambda(n) = 2(1 - (1 - 2\gamma^2 n^{-\alpha})^{1/4}) = O(n^{-\alpha/4}) \tag{6.3}$$

so that γ appears as a non-centrality parameter as is typically the case in the local-to-unity asymptotic analysis. Note that as $n \to \infty$ we obtain the non-stationary OU process with $\lambda(n) = 0$. Large values of γ lower the persistence.

The purpose of the simulation is two-fold. First we want to assess the efficiency gains of the optimal 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^{unc-hc}. This will allow us to appraise the difference between conditional and

[6]A recent extension to block local-to-unity asymptotics by Phillips, Moon, and Xiao (2001) has some resemblance with our analysis, although we focus here on integrated volatility.

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^{f-unc}, followed by the heteroskedasticity correction embedded in $\omega^{f-unc-hc}$.

We consider a 1,000 replications, each consisting of 500 and 1,000 "days.' We report the results for a total of three different continuous-time models along with n = 288, 144 and 24 corresponding to the use of five-minute, ten-minute and hourly returns in a 24-hour financial market. The class of models we simulate are based on Andersen, Bollerslev, and Meddahi (2005) and consist of:

$$\begin{aligned} d\log\ S_t &= \mu dt + \sigma_t dW_t \\ &= \sigma_t[\rho_1 dW_{1t} - \rho_2 dW_{2t} + \sqrt{1-\rho_1^2-\rho_2^2} dW_{3t}] \end{aligned} \tag{6.4}$$

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

$$d\log\ S_t = \sigma_t dW_{3t} \tag{6.5}$$

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

$$d\sigma_t^2 = .035(.636 - \sigma_t^2)dt + .144\sigma_t^2 dW_{1t} \tag{6.6}$$

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

$$\begin{aligned} d\sigma_{1t}^2 &= .5708(.3257 - \sigma_{1t}^2)dt + .2286\sigma_{1t}^2 dW_{1t} \\ d\sigma_{2t}^2 &= .0757(.1786 - \sigma_{2t}^2)dt + .1096\sigma_{2t}^2 dW_{2t} \end{aligned} \tag{6.7}$$

All of the above models satisfy the regularity conditions of the Jacod (1994) and Barndorff-Nielsen and Shephard (2002) asymptotics.

We start by treating the one-minute quantities as the "truth", hence they provide us with a benchmark for comparison. The results appear in

7 An Empirical Application

The following empirical results are based on GM (2000-2002) using 5min returns.

- Optimal weight for QV, No JUMP

$$\hat{\omega}_t^* = \frac{2Q_t}{2Q_t + 2\phi^2 Q_{t-1} + n[\overline{QV}_t - \overline{QV}_{t|t-1}]^2}$$

  where

  - n = number of obs. in a day
  - $QV_t = \sum_{j=1}^n r_{t,j}^2$
  - $Q_t = n/3 \sum_{j=1}^n r_{t,j}^4$

  Results: mean of weight=0.286, std of weight=0.302

- Optimal weight for QV, JUMP

$$\hat{\omega}_t^* = \frac{2Q_t}{2Q_t + 2\phi^2 Q_{t-1} + n[\overline{QV}_t - \overline{QV}_{t|t-1}]^2}$$

  where

  - n = number of obs. in a day
  - $QV_t = \sum_{j=1}^n r_{t,j}^2$
  - $Q_t = n(\pi/2)^2 \sum_{j=4}^n |r_{t,j} r_{t,j-1} r_{t,j-2} r_{t,j-3}|$

  Results: mean of weight=0.227, std of weight=0.284

- Optimal weight for BPV

$$\hat{\omega}_t^* = \frac{\nu Q_t}{\nu Q_t + \nu\phi^2 Q_{t-1} + n[\overline{BPV}_t - \overline{BPV}_{t|t-1}]^2}$$

  where

- n = number of obs. in a day
- $\nu = 0.609$
- $BPV_t = (\pi/2)\sum_{j=2}^{n}|r_{t,j}r_{t,j-1}|$
- $Q_t = n(\pi/2)^2\sum_{j=4}^{n}|r_{t,j}r_{t,j-1}r_{t,j-2}r_{t,j-3}|$

Results: mean of weight=0.153, std of weight=0.223

8 Conclusions

Technical Appendices

A Proof of Proposition 3.1

It should first be noted that:

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

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 (4.15) 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).

B Proof of Proposition 3.2

First, consider equation (3.3), namely:

$$\sigma_t^2 \equiv \frac{1}{n}\sum_{j=1}^{n}\sigma_{t,j}^2$$

Therefore:

$$Var\sigma_t^2 = \frac{1}{n^2}\sum_{i,j} Cov[\sigma_{t,i}^2, \sigma_{t,j}^2] \tag{B.2}$$

Recall also from equation (3.4) that $\hat{\sigma}_t^2 = \sum_{j=1}^{n}[r_{t,j}^n]^2$, and therefore:

$$Var\hat{\sigma}_t^2 = \sum_{i,j} Cov[r_{t,i}^2, r_{t,j}^2] \tag{B.3}$$

where: $Cov[r_{t,i}^2, r_{t,j}^2] = \frac{1}{n^2}E[\sigma_{t,i}^2\sigma_{t,j}^2]E[\varepsilon_{t,i}^2\varepsilon_{t,j}^2]$ - $\frac{1}{n^2}E[\sigma_{t,i}^2]E[\sigma_{t,j}^2]E[\varepsilon_{t,i}^2]E[\varepsilon_{t,j}^2]$, and therefore:

$$Cov[r_{t,i}^2, r_{t,j}^2] = \begin{cases} Cov[\sigma_{t,i}^2, \sigma_{t,j}^2] & i \neq j \\ \\ Var(\sigma_{t,i}^2) + 2E[\sigma_{t,i}^4] & i = j \end{cases}$$

Hence,

$$Var\hat{\sigma}_t^2 = Var\sigma_t^2 + -\frac{2}{n^2}\sum_j E[\sigma_{t,j}^4]$$

or by (3.15):

$$Var\hat{\sigma}_t^2 = Var\sigma_t^2 + \frac{2}{n}E[\sigma_t^4] \tag{B.4}$$

This yields the denominator of φ. To obtain an expression for the numerator, we derive:

$$\begin{aligned} Cov[\hat{\sigma}_t^2, \hat{\sigma}_{t-1}^2] &= Cov[\sum_j r_{t,j}^2, \sum_i r_{t-1,i}^2] \\ &= \frac{1}{n^2}\sum_{i,j}[E[\sigma_{t,i}^2\sigma_{t-1,j}^2\varepsilon_{t,i}^2\varepsilon_{t-1,j}^2] - E[\sigma_{t,i}^2\varepsilon_{t,i}^2]E[\sigma_{t-1,j}^2\varepsilon_{t-1,j}^2]] \\ &= \frac{1}{n^2}\sum_{i,j} Cov(\sigma_{t,i}^2, \sigma_{t-1,j}^2) \\ &= Cov(\sigma_t^2, \sigma_{t-1}^2) \end{aligned} \tag{B.5}$$

Therefore:

$$\varphi = \frac{Cov(\sigma_t^2, \sigma_{t-1}^2)}{Var\sigma_t^2 + (2/n)(E[\sigma_t^{[4]}])} \tag{B.6}$$

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).

C A Comparison of Two Estimators

Given the MLE estimator appearing in (3.20) we know that:

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

Hence, the expectation of the above the expectation of the above ratio is $(2n+n^2)/n^2 = (1+2/n)$. Hence, an unbiased estimator of $\sigma^{[4]} = m/n \sum_{t=1}^{m/n} \sigma^4_{t,[i]}$, is defined in equation (3.20). We can rewrite this as,

$$\begin{aligned}\hat{\sigma}^{[4]} &= \frac{m}{n}\sum_{t=1}^{m/n} \frac{\hat{\sigma}^4_{t,[i]}}{1+2/n} \\ &= \frac{m}{n(1+2/m)} \sum_i \frac{n^2}{m^2}(\sum r^2_{t,j})^2\end{aligned}$$

Which we can rewrite as:

$$\hat{\sigma}^{[4]} = \frac{n}{m+2}\sum_{i=1}^{n/m}[\sum_{j=n(t-1)+1}^{j=m} r^2_{t,j}]^2 \tag{C.7}$$

The above estimator can be compared with the naive estimator appearing in (**??**). To do so we need to derive the conditional variance of $\hat{\sigma}^{[4]}$. Note that we can rewrite the estimator (3.20) as:

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

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

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

The above results yield:

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

We now turn our attention to the naive estimator appearing in (**??**), which we rewrite as in equation (3.23):

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

Therefore,

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

From these results we can deduce that there will be efficiency improvements provided that $(m+3)/(m+2) < 4/3$, or $3m+9 < 4m+8$, which implies $m > 1$.

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

$$\begin{aligned} \hat{\sigma}^4_{t,[i]} &= \sigma^4_{t,[i]} \frac{[\chi^2_i(m)]^2}{m^2} \\ &= \sigma^4_{t,[i]} \varepsilon^2_i \end{aligned}$$

Hence,

$$\hat{\sigma}^4_t = \frac{n}{n(n+2/n)} \sum_{i=1}^{m/n} \sigma^4_{t,[i]} \varepsilon^2_i \tag{C.8}$$

Therefore the unconditional variance of $\hat{\sigma}^4_t$ can be written as:

$$Var\hat{\sigma}^4_t = \frac{n^2}{n^2(n+2/n)^2} \sum_{i=1}^{m/n} Var[\sigma^4_{t,[i]}](E[\varepsilon^2_i])^2 + 2(Var\varepsilon^2_i)E[\sigma^8_{t,[i]}] \tag{C.9}$$

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,

$$\begin{aligned} Var\hat{\sigma}_t^4 &= \frac{m^2}{n^2}\sum_{i=1}^{m/n} Var[\sigma_{t,[i]}^4] + 2\frac{8m(m+3)}{m^2(m+2)}E[\sigma_{t,[i]}^8] \\ &= Var[\sigma_t^4] + 2\frac{8m(m+3)}{m^2(m+2)}E[\sigma_{t,[i]}^8] \end{aligned} \tag{C.10}$$

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, 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.

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2002): "Parametric and Non-Parametric Volatility Measurement," Handbook of Financial Econometrics, forthcoming.

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. 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 (2006): "Regular and Modified Kernel-Based Estimators of Integrated Variance: The Case with Independent Noise," Discussion paper, Department of Mathematical Sciences, University of Aarhus.

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.

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.

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.

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.

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 (2006): "Jumps in Financial Markets: A New Nonparametric Test and Jump Dynamics," Discussion paper, Georgia Institute of Technology and The University of Chicago, to appear in *Review of Financial Studies*.

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.

Meddahi, N., and E. Renault (2004): "Temporal Aggregation of Volatility Models," *Journal of Econometrics*, 119, 355–379.

Mykland, P., and L. Zhang (2007): "Inference for continuous semimartingales observed at high frequency: A general approach (revised version)," Discussion paper, in revision for *Econometrica*.

Mykland, P. A. (2006): "A Gaussian Calculus for Inference from High Frequency Data," Discussion paper, University of Chicago.

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

Phillips, P. (1987): "Towards a unified asymptotic theory for autoregression," *Biometrika*, 74, 535547.

Phillips, P., R. Moon, and Z. Xiao (2001): "How to estimate autoregressive roots near unity," *Econometric Theory*, 17, 2969.

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.

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, The University of Chicago, Department of Statistics.

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

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

Table 1: MSE Improvements, GARCH Diffusion Model

AR(1) Prediction, Sample size=500 days

	MSE raw	Impr: ω^*		Impr: ω^{rth}		Impr:ω^{unc}		Impr:ω^{unc-hc}	
	Mean	Mean	Var	Mean	Var	Mean	Var	Mean	Var
1 Min	0.000698	0.968657	0.000569	2.933233	0.164681	0.938143	0.000324	0.938207	0.000325
5 Min	0.003468	0.688474	0.001149	0.986316	0.013489	0.835003	0.000482	0.812633	0.000994
10 Min	0.006872	0.617635	0.001208	0.745121	0.005736	0.790329	0.000432	0.749763	0.001184
1 Hour	0.040485	0.552005	0.002531	0.550450	0.001761	0.726488	0.000575	0.664082	0.002530

AR(1) Prediction, Sample size=1000 days

	MSE raw	Impr: ω^*		Impr: ω^{rth}		Impr:ω^{unc}		Impr:ω^{unc-hc}	
	Mean	Mean	Var	Mean	Var	Mean	Var	Mean	Var
1 Min	0.000740	0.968851	0.000323	2.930422	0.093108	0.938663	0.000181	0.938735	0.000182
5 Min	0.003708	0.689292	0.000729	0.987504	0.007240	0.836202	0.000278	0.814704	0.000583
10 Min	0.007343	0.618920	0.000732	0.745429	0.003442	0.791686	0.000247	0.751894	0.000711
1 Hour	0.043333	0.552880	0.001553	0.550193	0.001067	0.729105	0.000362	0.668533	0.001619

Table 2: Comparison of weights, GARCH Diffusion Model

Sample size=500 days

	ω^*		ω^{unc}			ω^{unc-hc}		
	Mean	Var	Mean	Var	meanvar	Mean	Var	meanvar
1 Min	0.149728	0.023604	0.040616	0.000014	0.000000	0.040584	0.000014	0.000000
5 Min	0.240160	0.026001	0.104851	0.000067	0.000000	0.124912	0.000086	0.000035
10 Min	0.272482	0.026180	0.131608	0.000092	0.000000	0.169400	0.000121	0.000185
1 Hour	0.318790	0.037506	0.174333	0.000139	0.000000	0.248265	0.000327	0.003698

Sample size=1000 days

	ω^*		ω^{unc}			ω^{unc-hc}		
	Mean	Var	Mean	Var	meanvar	Mean	Var	meanvar
1 Min	0.149602	0.023466	0.040369	0.000008	0.000000	0.040327	0.000008	0.000000
5 Min	0.240637	0.025864	0.104163	0.000041	0.000000	0.123741	0.000050	0.000034
10 Min	0.273390	0.025979	0.130415	0.000057	0.000000	0.167491	0.000067	0.000176
1 Hour	0.321076	0.037370	0.172181	0.000096	0.000000	0.243963	0.000214	0.003492

Table 3: MSE, GARCH Diffusion Model

AR(1) Prediction, Sample size=500 days

	MSE raw		MSE: ω^*		MSE: ω^{rth}		MSE:ω^{unc}		MSE:ω^{unc-hc}	
	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var
1 Min	0.000698	0.000000	0.000676	0.000000	0.002047	0.000002	0.000656	0.000000	0.000656	0.000000
5 Min	0.003468	0.000005	0.002395	0.000002	0.003436	0.000006	0.002905	0.000003	0.002829	0.000003
10 Min	0.006872	0.000018	0.004260	0.000007	0.005129	0.000011	0.005447	0.000012	0.005171	0.000011
1 Hour	0.040485	0.000617	0.022478	0.000209	0.022283	0.000189	0.029580	0.000347	0.027135	0.000308

AR(1) Prediction, Sample size=1000 days

	MSE raw		MSE: ω^*		MSE: ω^{rth}		MSE:ω^{unc}		MSE:ω^{unc-hc}	
	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var
1 Min	0.000740	0.000000	0.000717	0.000000	0.002172	0.000001	0.000695	0.000000	0.000696	0.000000
5 Min	0.003708	0.000003	0.002560	0.000002	0.003659	0.000003	0.003105	0.000002	0.003025	0.000002
10 Min	0.007343	0.000011	0.004555	0.000005	0.005472	0.000007	0.005822	0.000007	0.005531	0.000007
1 Hour	0.043333	0.000414	0.024050	0.000145	0.023848	0.000131	0.031725	0.000239	0.029156	0.000216

Table 4: Bias2+Var, GARCH Diffusion Model

AR(1) Prediction, Sample size=500 days

	MSE raw		MSE: ω^*		MSE: ω^{rth}		MSE:ω^{unc}		MSE:ω^{unc-hc}	
	Bias2	Var	Bias2	Var	Bias2	Var	Bias2	Var	Bias2	Var
1 Min	0.000002	0.000698	0.000002	0.000676	0.000003	0.002049	0.000002	0.000655	0.000002	0.000655
5 Min	0.000012	0.003463	0.000022	0.002378	0.000015	0.003428	0.000012	0.002899	0.000014	0.002820
10 Min	0.000033	0.006853	0.000076	0.004193	0.000038	0.005101	0.000033	0.005425	0.000046	0.005135
1 Hour	0.000727	0.039838	0.002222	0.020297	0.000756	0.021570	0.000727	0.028912	0.001402	0.025785

AR(1) Prediction, Sample size=1000 days

	MSE raw		MSE: ω^*		MSE: ω^{rth}		MSE:ω^{unc}		MSE:ω^{unc-hc}	
	Bias2	Var	Bias2	Var	Bias2	Var	Bias2	Var	Bias2	Var
1 Min	0.000001	0.000740	0.000001	0.000717	0.000001	0.002172	0.000001	0.000695	0.000001	0.000695
5 Min	0.000008	0.003703	0.000019	0.002543	0.000010	0.003653	0.000008	0.003099	0.000011	0.003017
10 Min	0.000026	0.007325	0.000072	0.004488	0.000029	0.005449	0.000026	0.005802	0.000040	0.005497
1 Hour	0.000710	0.042666	0.002263	0.021809	0.000726	0.023145	0.000710	0.031046	0.001415	0.027769

Table 5: Forecasting scheme, GARCH Diffusion Model

	AR(1) with IV	AR(1) with RV	IV∼RV	IV∼Corrected RV
		Sample size: 500 days		
		MSE		
1 Min	0.007	0.008	0.007	0.007
5 Min	0.007	0.013	0.010	0.010
10 Min	0.007	0.019	0.013	0.012
1 Hour	0.007	0.073	0.036	0.031
		R2		
1 Min	0.935	0.922	0.929	0.927
5 Min	0.935	0.874	0.905	0.905
10 Min	0.935	0.820	0.876	0.881
1 Hour	0.935	0.470	0.664	0.700
		Sample size: 1000 days		
		MSE		
1 Min	0.007	0.009	0.008	0.008
5 Min	0.007	0.015	0.011	0.011
10 Min	0.007	0.022	0.014	0.013
1 Hour	0.007	0.082	0.041	0.036
		R2		
1 Min	0.945	0.935	0.940	0.939
5 Min	0.945	0.893	0.919	0.920
10 Min	0.945	0.845	0.894	0.900
1 Hour	0.945	0.517	0.700	0.736

Table 6: MSE Improvements, Two-factor diffusion Model

AR(1) Prediction, Sample size=500 days

	MSE raw	Impr: ω^*		Impr: ω^{rth}		Impr:ω^{unc}		Impr:ω^{unc-hc}	
	Mean	Mean	Var	Mean	Var	Mean	Var	Mean	Var
1 Min	0.000385	0.989081	0.000259	6.335047	0.567550	0.967679	0.000098	0.967674	0.000099
5 Min	0.001921	0.741426	0.000693	1.665778	0.028773	0.886338	0.000244	0.876034	0.000409
10 Min	0.003829	0.656901	0.000800	1.081521	0.009140	0.831873	0.000267	0.806082	0.000638
1 Hour	0.022474	0.531366	0.001562	0.605291	0.001340	0.708880	0.000253	0.626715	0.001404

AR(1) Prediction, Sample size=1000 days

	MSE raw	Impr: ω^*		Impr: ω^{rth}		Impr:ω^{unc}		Impr:ω^{unc-hc}	
	Mean	Mean	Var	Mean	Var	Mean	Var	Mean	Var
1 Min	0.000389	0.989200	0.000132	6.255583	0.300284	0.967938	0.000052	0.967936	0.000053
5 Min	0.001941	0.740910	0.000353	1.650102	0.015274	0.886331	0.000129	0.876288	0.000216
10 Min	0.003875	0.657091	0.000420	1.073845	0.004683	0.831778	0.000139	0.806360	0.000330
1 Hour	0.022823	0.530654	0.000803	0.603472	0.000691	0.708773	0.000132	0.626347	0.000762

Table 7: Comparison of weights, Two-factor diffusion Model

Sample size=500 days

	ω^*		ω^{unc}			ω^{unc-hc}		
	Mean	Var	Mean	Var	meanvar	Mean	Var	meanvar
1 Min	0.108224	0.019962	0.021156	0.000003	0.000000	0.021120	0.000003	0.000000
5 Min	0.195840	0.026460	0.073265	0.000023	0.000000	0.082200	0.000032	0.000014
10 Min	0.239036	0.027846	0.106603	0.000035	0.000000	0.129582	0.000058	0.000107
1 Hour	0.329995	0.038291	0.182154	0.000073	0.000000	0.267595	0.000146	0.005350

Sample size=1000 days

	ω^*		ω^{unc}			ω^{unc-hc}		
	Mean	Var	Mean	Var	meanvar	Mean	Var	meanvar
1 Min	0.108309	0.019884	0.021260	0.000002	0.000000	0.021223	0.000002	0.000000
5 Min	0.195985	0.026357	0.073471	0.000012	0.000000	0.082386	0.000018	0.000014
10 Min	0.239039	0.027730	0.106674	0.000019	0.000000	0.129625	0.000032	0.000107
1 Hour	0.330089	0.038269	0.181439	0.000036	0.000000	0.266882	0.000078	0.005308

Table 8: MSE, Two-factor diffusion Model

AR(1) Prediction, Sample size=500 days

	MSE raw		MSE: ω^*		MSE: ω^{rth}		MSE:ω^{unc}		MSE:ω^{unc-hc}	
	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var
1 Min	0.000385	0.000000	0.000381	0.000000	0.002413	0.000000	0.000373	0.000000	0.000373	0.000000
5 Min	0.001921	0.000000	0.001422	0.000000	0.003168	0.000000	0.001701	0.000000	0.001681	0.000000
10 Min	0.003829	0.000000	0.002512	0.000000	0.004109	0.000000	0.003184	0.000000	0.003083	0.000000
1 Hour	0.022474	0.000010	0.011919	0.000003	0.013568	0.000003	0.015938	0.000005	0.014070	0.000004

AR(1) Prediction, Sample size=1000 days

	MSE raw		MSE: ω^*		MSE: ω^{rth}		MSE:ω^{unc}		MSE:ω^{unc-hc}	
	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var
1 Min	0.000389	0.000000	0.000385	0.000000	0.002418	0.000000	0.000376	0.000000	0.000376	0.000000
5 Min	0.001941	0.000000	0.001438	0.000000	0.003186	0.000000	0.001720	0.000000	0.001700	0.000000
10 Min	0.003875	0.000000	0.002545	0.000000	0.004144	0.000000	0.003222	0.000000	0.003123	0.000000
1 Hour	0.022823	0.000005	0.012100	0.000002	0.013751	0.000002	0.016180	0.000003	0.014287	0.000002

Table 9: Bias2+Var, Two-factor diffusion Model

AR(1) Prediction, Sample size=500 days

	MSE raw		MSE: ω^*		MSE: ω^{rth}		MSE:ω^{unc}		MSE:ω^{unc-hc}	
	Bias2	Var	Bias2	Var	Bias2	Var	Bias2	Var	Bias2	Var
1 Min	0.000001	0.000385	0.000001	0.000380	0.000001	0.002416	0.000001	0.000373	0.000001	0.000373
5 Min	0.000007	0.001918	0.000016	0.001409	0.000008	0.003166	0.000007	0.001698	0.000008	0.001676
10 Min	0.000019	0.003817	0.000058	0.002459	0.000023	0.004095	0.000019	0.003170	0.000027	0.003062
1 Hour	0.000476	0.022043	0.001575	0.010365	0.000492	0.013102	0.000476	0.015494	0.000990	0.013106

AR(1) Prediction, Sample size=1000 days

	MSE raw		MSE: ω^*		MSE: ω^{rth}		MSE:ω^{unc}		MSE:ω^{unc-hc}	
	Bias2	Var	Bias2	Var	Bias2	Var	Bias2	Var	Bias2	Var
1 Min	0.000001	0.000389	0.000001	0.000384	0.000001	0.002419	0.000001	0.000376	0.000001	0.000376
5 Min	0.000005	0.001938	0.000015	0.001424	0.000006	0.003184	0.000005	0.001717	0.000006	0.001696
10 Min	0.000016	0.003863	0.000056	0.002491	0.000017	0.004130	0.000016	0.003209	0.000024	0.003102
1 Hour	0.000455	0.022391	0.001573	0.010537	0.000463	0.013302	0.000455	0.015741	0.000978	0.013322

Table 10: Forecasting scheme, Two-factor diffusion Model

	AR(1) with IV	AR(1) with RV	IV∼RV	IV∼Corrected RV
		Sample size: 500 days		
		MSE		
1 Min	0.008	0.009	0.008	0.008
5 Min	0.008	0.011	0.009	0.009
10 Min	0.008	0.014	0.010	0.010
1 Hour	0.008	0.037	0.016	0.016
		R2		
1 Min	0.670	0.650	0.660	0.656
5 Min	0.670	0.577	0.624	0.618
10 Min	0.670	0.503	0.583	0.581
1 Hour	0.670	0.177	0.350	0.363
		Sample size: 1000 days		
		MSE		
1 Min	0.008	0.009	0.008	0.008
5 Min	0.008	0.011	0.009	0.009
10 Min	0.008	0.014	0.010	0.010
1 Hour	0.008	0.038	0.016	0.016
		R2		
1 Min	0.679	0.658	0.668	0.665
5 Min	0.679	0.585	0.631	0.627
10 Min	0.679	0.510	0.590	0.590
1 Hour	0.679	0.179	0.353	0.368

Table 11: Forecasting scheme, Two-factor diffusion Model with leverage effect

	AR(1) with IV	AR(1) with RV	IV∼RV	IV∼Corrected RV
		Sample size: 500 days		
		MSE		
1 Min	0.008	0.009	0.008	0.008
5 Min	0.008	0.011	0.009	0.009
10 Min	0.008	0.014	0.010	0.010
1 Hour	0.008	0.038	0.016	0.015
		R2		
1 Min	0.670	0.650	0.661	0.657
5 Min	0.670	0.578	0.625	0.620
10 Min	0.670	0.503	0.586	0.583
1 Hour	0.670	0.179	0.363	0.376
		Sample size: 1000 days		
		MSE		
1 Min	0.008	0.009	0.008	0.008
5 Min	0.008	0.011	0.009	0.009
10 Min	0.008	0.014	0.010	0.010
1 Hour	0.008	0.038	0.016	0.016
		R2		
1 Min	0.679	0.658	0.669	0.666
5 Min	0.679	0.585	0.633	0.629
10 Min	0.679	0.509	0.593	0.592
1 Hour	0.679	0.181	0.366	0.381

Table 12: Forecasting scheme, Two-factor diffusion Model with leverage and drfit

	AR(1) with IV	AR(1) with RV	IV∼RV	IV∼Corrected RV
		Sample size: 500 days		
		MSE		
1 Min	0.008	0.009	0.008	0.008
5 Min	0.008	0.011	0.009	0.009
10 Min	0.008	0.014	0.010	0.010
1 Hour	0.008	0.038	0.016	0.015
		R2		
1 Min	0.670	0.650	0.661	0.657
5 Min	0.670	0.578	0.626	0.620
10 Min	0.670	0.503	0.586	0.584
1 Hour	0.670	0.180	0.365	0.378
		Sample size: 1000 days		
		MSE		
1 Min	0.008	0.009	0.008	0.008
5 Min	0.008	0.011	0.009	0.009
10 Min	0.008	0.014	0.010	0.010
1 Hour	0.008	0.039	0.016	0.016
		R2		
1 Min	0.679	0.658	0.669	0.666
5 Min	0.679	0.585	0.633	0.629
10 Min	0.679	0.509	0.593	0.592
1 Hour	0.679	0.182	0.368	0.383

Table 13: Forecasting scheme, GARCH Diffusion Model

	AR(1) with IV	AR(1) with RV	IV∼RV	IV∼Corrected RV
		Sample size: 500 days		
		MSE		
1 Min	0.007	0.008	0.007	0.007
5 Min	0.007	0.013	0.010	0.010
10 Min	0.007	0.019	0.013	0.012
1 Hour	0.007	0.073	0.036	0.031
		R2		
1 Min	0.935	0.922	0.929	0.927
5 Min	0.935	0.874	0.905	0.905
10 Min	0.935	0.820	0.876	0.881
1 Hour	0.935	0.470	0.664	0.700
		Sample size: 1000 days		
		MSE		
1 Min	0.007	0.009	0.008	0.008
5 Min	0.007	0.015	0.011	0.011
10 Min	0.007	0.022	0.014	0.013
1 Hour	0.007	0.082	0.041	0.036
		R2		
1 Min	0.945	0.935	0.940	0.939
5 Min	0.945	0.893	0.919	0.920
10 Min	0.945	0.845	0.894	0.900
1 Hour	0.945	0.517	0.700	0.736