Essays on Estimation and Inference for Volatility with High Frequency Data

Ilze Kalnina

A Dissertation
Presented to the Department of Economics
of London School of Economics and Political Science
in Candidacy for the Degree
of Doctor of Philosophy

June 2009

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without the prior written consent of the author.

I warrant that this authorization does not, to the best of my belief, infringe the rights of any third party.

Abstract

Volatility is a measure of risk, and as such it is crucial for finance. But volatility is not observable, which is why estimation and inference for it are important. Large high frequency data sets have the potential to increase the precision of volatility estimates. However, this data is also known to be contaminated by market microstructure frictions, such as bid-ask spread, which pose a challenge to estimation of volatility.

The first chapter, joint with Oliver Linton, proposes an econometric model that captures the effects of market microstructure on a latent price process. In particular, this model allows for correlation between the measurement error and the return process and allows the measurement error process to have diurnal heteroskedasticity. A modification of the TSRV estimator of quadratic variation is proposed and asymptotic distribution derived.

Financial econometrics continues to make progress in developing more robust and efficient estimators of volatility. But for some estimators, the asymptotic variance is hard to derive or may take a complicated form and be difficult to estimate. To tackle these problems, the second chapter develops an automated method of inference that does not rely on the exact form of the asymptotic variance. The need for a new approach is motivated by the failure of traditional bootstrap and subsampling variance estimators with high frequency data, which is explained in the paper. The main contribution is to propose a novel way of conducting inference for an important general class of estimators that includes many estimators of integrated volatility. A subsampling scheme is introduced that consistently estimates the asymptotic variance for an estimator, thereby facilitating inference and the construction of valid confidence intervals.

The third chapter shows how the multivariate version of the subsampling method of Chapter 2 can be used to study the question of time variability in equity betas.

Acknowledgements

I am deeply indebted to my supervisor Oliver B. Linton for his help, encouragement, and guidance throughout my PhD studies. I am grateful to Peter C. B. Phillips for hosting me at the Cowles Foundation and all the support and inspiration I have received there.

This research was supported by the Cowles Foundation, ESRC, and the Leverhulme Foundation.

Contents

Chapter 1

Estimation of Quadratic Variation in the Presence of Diurnal and Heteroscedastic Measurement Error

1.1 Introduction

It has been widely recognized that using very high frequency data requires taking into account the effect of market microstructure (MS) noise. We are interested in the estimation of the quadratic variation of a latent price in the case where the observed log-price Y is a sum of the latent log-price X that evolves in continuous time and an error u that captures the effect of MS noise.

There is by now a large literature that uses realized variance as a nonparametric measure of volatility. The justification is that in the absence of market microstructure noise it is a consistent estimator of the quadratic variation as the time between observations goes to zero. For a literature review, see Barndorff-Nielsen and Shephard

(2007). In practice, ignoring microstructure noise seems to work well for frequencies below 10 minutes. For higher frequencies realized variance is not robust, as has been evidenced in the so-called 'volatility signature plots', see, e.g. Andersen et al. (2000).

The additive measurement error model where u is independent of X and i.i.d. over time was first introduced by Zhou (1996). The usual realized volatility estimator is inconsistent under this assumption. The first consistent estimator of quadratic variation of the latent price in the presence of MS noise was proposed by Zhang, Mykland, and Aït-Sahalia (2005a) who introduced the Two Scales Realized Volatility (TSRV) estimator, and derived the appropriate central limit theory. TSRV estimates the quadratic variation using a combination of realized variances computed on two different time scales, performing an additive bias correction. It has a rate of convergence $n^{-1/6}$. Zhang (2004) introduced the more complicated Multiple Scales Realized Volatility (MSRV) estimator that combines multiple ($\sim n^{1/2}$) time scales, which has a convergence rate of $n^{-1/4}$. This is known to be the optimal rate for this problem. Both papers assumed that the MS noise was i.i.d. and independent of the latent price. This assumption, according to an empirical analysis of Hansen and Lunde (2006), "seems to be reasonable when intraday returns are sampled every 15 ticks or so". Further studies have tried to relax this assumption to allow modelling of even higher frequency returns. Aït-Sahalia, Mykland and Zhang (2006a) modify TSRV and MSRV estimators and achieve consistency in the presence of serially correlated MS noise. Another class of consistent estimators of the quadratic variation was proposed by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006). They introduce realized kernels, a general class of estimators that extends the unbiased but inconsistent estimator of Zhou (1996), and is based on a general weighting of realized autocovariances as well as realized variances. They show that realized kernels can be designed to be consistent and derive the central limit theory. They show that for particular choices of weight functions they can be asymptotically equivalent to TSRV and MSRV estimators, or even more efficient. Apart from the benchmark setup where the noise is i.i.d. and independent from the latent price Barndorff-Nielsen et al. (2006) have two additional sections, one allowing for AR(1) structure in the noise, another with an additional endogenous term albeit one that is asymptotically degenerate.

We generalize the standard additive noise model (where the noise is i.i.d. and independent from the latent price) in three directions. The first generalization is allowing for (asymptotically non-degenerate) correlation between MS noise and the latent returns. This is motivated by a paper of Hansen and Lunde (2006), where, for very high frequencies: "the key result is the overwhelming evidence against the independent noise assumption. This finding is quite robust to the choice of sampling method (calendar-time or tick-time) and the type of price data (transaction prices or quotation prices)".[1]

Another generalization concerns the magnitude of the MS noise. All of the papers above, like most of related literature, assume that the variance of the MS noise is constant and does not change depending on the time interval between trades. We call this a large noise assumption. We explicitly model the magnitude of the MS noise via a parameter α, where the $\alpha = 0$ case corresponds to the benchmark case of large noise. We allow also $\alpha > 0$ in which case the noise is "small" and specifically the variance of the noise shrinks to zero with the sample size n. The rate of convergence of our estimator depends on the magnitude of the noise, and can be from $n^{-1/6}$ to $n^{-1/3}$, where $n^{-1/6}$ is the rate of convergence corresponding to the "big" noise case when $\alpha = 0$.

How could the size of the noise "depend" on the sample size? We give a fuller discussion of this issue below, but we note here two arguments. First, there is a

[1]By "independent noise" Hansen and Lunde (2006) mean the combination of the i.i.d. assumption and the assumption that the noise is independent from the latent price. Our paper proposes to relax the second assumption. As to the first assumption, we do not allow for serial correlation in the noise. At the same time, we only impose approximate stationarity compared to Hansen and Lunde (2006) since we allow for intraday heteroscedasticity of the noise.

negative relationship between the bid-ask spread (an important component of the MS noise for transaction data) and a number of (other) liquidity measures, including number of transactions during the day. This negative relationship is a stylized fact from the market microstructure literature. See, for example, Copeland and Galai (1983) and McInish and Wood (1992). Also, Awartani, Corradi and Distaso (2004) write that "an alternative model of economic interest [to the standard additive noise model] would be one in which the microstructure noise variance is positively correlated with the time interval". This is in principle a testable hypothesis. Using Dow Jones Industrial Average data, the authors test for and reject the hypothesis of constant variance of the MS noise across frequencies.

The third feature of our model is that we allow the MS noise to exhibit diurnal heteroscedasticity. This is motivated by the stylized fact in market microstructure literature that intradaily spreads and intradaily stock price volatility are described typically by a U-shape (or reverse J-shape). See Andersen and Bollerslev (1997), Gerety and Mulherin (1994), Harris (1986), Kleidon and Werner (1996), Lockwood and Linn (1990), and McInish and Wood (1992). Allowing for diurnal heteroscedasticity in our model has the effect that the original TSRV estimator may not be consistent because of end effects. In some cases, instead of estimating the quadratic variation, it would be estimating some function of the noise. We propose a modification of the TSRV estimator that is consistent, without introducing new parameters to be chosen. Our model is not meant to be definitive and can be generalized in a number of ways.

The structure of the paper is as follows. Section 2 introduces the model. Section 3 describes the estimator. Section 4 gives the main result and the intuition behind it. Section 5 investigates the numerical properties of the estimator in a set of simulation experiments. Section 6 illustrates the ideas with an empirical study of IBM transaction prices. Section 7 concludes. We use $\Longrightarrow$ to denote convergence in distribution.

1.2 The Model

Suppose that the latent (log) price process $\{X_t, t \in [0,T]\}$ is a Brownian semimartingale solving the stochastic differential equation

$$dX_t = \mu_t dt + \sigma_t dW_t, \tag{1.1}$$

where W_t is standard Brownian motion, μ_t is a locally bounded predictable drift function, and σ_t a càdlàg volatility function; both are independent of the process $\{W_t, t \in [0,T]\}$. The (no leverage) assumption of $\{\sigma_t, \mu_t, t \in [0,T\}$ being independent of $\{W_t, t \in [0,T]\}$, though reasonable for exchange rate data, is unrealistic for stock price data. However, it is frequently used and makes the theoretical analysis more tractable. The simulation results suggest that this assumption does not change the result. Furthermore, in many other contexts the presence of leverage does not affect the limiting distributions, see Barndorff-Nielsen and Shephard (2002).

The additive noise model says that the noisy price Y is observed at times $t_1, \ldots, t_n$ on some fixed domain $[0,T]$

$$Y_{t_i} = X_{t_i} + u_{t_i}, \tag{1.2}$$

where u_{t_i} is a random variable representing measurement error. Without loss of much generality we are going to restrict attention to the case of equidistant observations with $T = 1$. This type of model was first introduced by Zhou (1996) who assumed that u_{t_i} is i.i.d. over i and independent of $\{X_t, t \in [0,1]\}$. In this case the signal to noise ratio for returns decreases with sample size, i.e., $\text{var}(\Delta X_{t_i})/\text{var}(\Delta u_{t_i}) \to 0$ as $n \to \infty$, and at a specific rate such that $\lim_{n\to\infty} n\text{var}(\Delta X_{t_i})/\text{var}(\Delta u_{t_i}) < \infty$, which implies inconsistency of realized volatility. We are going to modify the properties of the process $\{u_{t_i}\}$ and its relation to $\{X_t, t \in [0,1]\}$.

We would like to capture the idea that the measurement error can be small. This can be addressed by adopting a model $u_{t_i} = \sigma_\epsilon \epsilon_{t_i}$, where ϵ_{t_i} is an i.i.d. sequence with mean zero and variance one, and σ_ϵ is a parameter such that $\sigma_\epsilon \to 0$. Many authors have found small σ_ϵ in practice. As usual one wants to make inferences about data drawn from the true probability measure of the data where both n is finite and $\sigma_\epsilon > 0$ by working with a limiting case that is more tractable. In this case there are a variety of limits that one could take. Bandi and Russell (2006a) for example calculate the exact MSE of the statistic of interest, and then in equation (24) implicitly take $\sigma_\epsilon \to 0$ followed by $n \to \infty$. We instead take the sound and well established practice in econometrics of taking pathwise limits, that is we let $\sigma_\epsilon = \sigma_\epsilon(n)$ and then let $n \to \infty$. Such a limit with "small" noise has been used before to derive Edgeworth approximations (Zhang et al., 2005b), to calculate optimal sampling frequency of inconsistent estimator for QV_x (Zhang et al., 2005a, eqn. 53), to estimate QV_x consistently when X follows a pure jump process and Y is observed fully and continuously (Large, 2007), and to estimate QV_x consistently in a pure rounding model (Li and Mykland, 2006; Rosenbaum, 2007). An example from MS modelling literature in microeconomics is Back and Baruch (2004) who show the link between the two key papers in asymmetric information modelling, Glosten and Milgrom (1985) and Kyle (1985) using a limit with small noise. In particular, they consider a limit of Glosten and Milgrom (1985) as the arrival rate of trades explodes (so the number of trades in any interval goes to infinity) and order size (and hence incremental information per trade) goes to zero, thus reaching the Kyle (1985) model as a limit. We are also mindful not to preclude the case where $\sigma_\epsilon(n)$ is "large" i.e., (in our framework) does not vanish with n, and our parameterization below allows us to do that.

We next present our model. We assume that

$$\begin{aligned} u_{t_i} &= v_{t_i} + \varepsilon_{t_i} \\ v_{t_i} &= \delta\gamma_n \left(W_{t_i} - W_{t_{i-1}}\right) \\ \varepsilon_{t_i} &= m\left(t_i\right) + n^{-\alpha/2}\omega\left(t_i\right)\epsilon_{t_i}, \;\; \alpha \in [0, 1/2) \end{aligned} \tag{1.3}$$

with ϵ_{t_i} i.i.d. mean zero and variance one and independent of the Gaussian process $\{W_t, t \in [0,1]\}$ with $E|\epsilon_{t_i}|^{4+\eta} < \infty$ for some $\eta > 0$. The functions m and ω are differentiable, nonstochastic functions of time. They are unknown as are the constants δ and α. The usual benchmark measurement error model with noise being i.i.d. and independent from the latent price has $\alpha = 0$, $\gamma_n = 0$ and $\omega(.)$ and $m(.)$ constant (see, e.g., Barndorff-Nielsen and Shephard (2002), Zhang et al. (2005a) and Bandi and Russell (2006b)).

The process for the latent log-price is motivated by the fundamental theory of asset prices, which states that, in a frictionless market, log-prices must obey a semimartingale; we are specializing to the Brownian semimartingale case (1.1). We want to model log-prices at very high frequency where frictions are important and observed prices do not follow a semimartingale. One way of partly reconciling the evidence in volatility signature plots of the price behavior in very high and moderate frequencies is to assume that observed prices can be decomposed as in (1.2). The first component X is a semi-martingale with finite quadratic variation, while the second component u is not a semi-martingale and has infinite quadratic variation. In particular, the increments in u are of larger magnitude than that of X, and this difference is the key in identifying the quadratic variation of X. We split the noise component u into an independent term ε that has been considered in the literature, and a 1-dependent endogenous part v, which is correlated with X due to being driven by the same Brownian motion. At the same time, v preserves the features of not being a semimartingale and having infinite quadratic variation, the main motivation of the way ε is modelled.

There are three key parts to our model: the correlation between u and X, the relative magnitudes of u and X, and the heterogeneity of u. We have $E[u_{t_i}] = m(t_i)$ and $\text{var}[u_{t_i}] = \delta^2\gamma_n^2(t_i - t_{i-1}) + 2n^{-\alpha}\sigma_\epsilon^2(i/n)$. To have the variance of both terms in u equal, we set $\gamma_n^2 = n^{1-\alpha}$. This seems like a reasonable restriction if both components are generated by the same mechanism. In this case, both of the measurement error terms are $O_p(n^{-\alpha})$. In our model the signal to noise ratio of returns varies with sample size in a way depending on α so that only $\lim_{n\to\infty} n^{1-\alpha}\text{var}(\Delta X_{t_i})/\text{var}(\Delta u_{t_i}) < \infty$. We exploit the fact that for consistency of the TSRV estimator, it is enough to assume that noise increments are of larger order of magnitude than the latent returns, and the usual stronger assumption $\lim_{n\to\infty} n\text{var}(\Delta X_{t_i})/\text{var}(\Delta u_{t_i}) < \infty$ is not necessary.

The process ε_{t_i} is a special case of the more general class of locally stationary processes of Dahlhaus (1997). The generalization to allowing time varying mean and variance in the measurement error allows one to capture diurnal variation in the measurement error process, which is likely to exist in calendar time. Nevertheless, the measurement error in prices is approximately stationary under our conditions, which seems reasonable.

The term v in u induces a correlation between latent returns and the change in the measurement error, which can be of either sign depending on δ. Correlation between u and X is plausible due to rounding effects, price stickiness, asymmetric information, or other reasons [Bandi and Russell (2006c), Hansen and Lunde (2006), Diebold (2006)].[2] In the special case that $\sigma_t = \sigma$ and $\omega(t_i) = \omega$, we find

$$\text{corr}(\Delta X_{t_i}, \Delta u_{t_i}) \simeq \frac{\delta}{\sqrt{[2\delta^2 + 2\omega^2]}}.$$

[2]In a recent survey of measurement error in microeconometrics models, Bound, Brown, and Mathiowetz (2001) emphasize 'mean-reverting' measurement error that is correlated with the signal.

In this case, the range of correlation is limited, although it is quite wide - one can obtain up to a correlation of $\pm 1/\sqrt{2}$ depending on the relative magnitudes of δ, ω.

An alternative model for endogenous noise has been developed by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006). In our notation, they have the endogenous noise part such that $\mathrm{var}(v_{t_i}) = O(1/n)$, and an i.i.d., independent from X part with $\mathrm{var}(\varepsilon_{t_i}) = O(1)$. They conclude robustness of their estimator to this type of endogeneity, with no change to the first order asymptotic properties compared to the case where $v_{t_i} = 0$.

The focus of this paper is on estimating increments in quadratic variation of the latent price process,[3] but estimation of parameters of the MS noise in our model is also of interest. We acknowledge that not all the parameters of our model are identifiable. In particular, the endogeneity parameter may not be identified unless one knows something about the distribution of ϵ and in particular that it is not Gaussian.[4] However, other parameters are identified. In Linton and Kalnina (2005) we provided a consistent estimator of α, see also Section 6 here for empirical implementation and discussion. Estimating the function $\omega(\tau)$ would allow us to measure the diurnal variation of the MS noise. In the benchmark measurement error model this is a constant $\omega(\tau) \equiv \omega$ that can be estimated consistently by $\sum_{i=1}^{n-1} \left(Y_{t_{i+1}} - Y_{t_i}\right)^2 / 2n$ (Bandi and Russell, 2006b; Barndorff-Nielsen et al., 2006; Zhang et al., 2005a). In

[3]There is a question about whether one should care about the latent price or the actual price. This has been raised elsewhere, see Zhang, Mykland, and Aït-Sahalia (2005). We stick with the usual practice here, acknowledging that the presence of correlation between the noise and efficient price makes this even more debatable, Aït-Sahalia, Mykland, and Zhang (2006b). Also, note that we are following the literature and estimating the quadratic variation of the latent log-price and not the latent price.

[4]Suppose that $X_{i+1} = X_i + (\sigma/\sqrt{n}) z_{i+1}$ and $Y_i = X_i + \rho z_i + \sigma_\epsilon \epsilon_i$, where z_i is standard normal and ε_i is i.i.d. with mean zero and variance one. Then $r_{i+1} = Y_{i+1} - Y_i = \left(\frac{\sigma}{\sqrt{n}} + \rho\right)' z_{i+1} - \rho z_i + \sigma_\epsilon \epsilon_{i+1} - \sigma_\epsilon \epsilon_i$. We have $\mathrm{var}[r_{i+1}] = 2\left(\rho^2 + \sigma_\epsilon^2\right) + \frac{2\rho\sigma}{\sqrt{n}} + \frac{\sigma^2}{n}$, $\mathrm{cov}[r_{i+1}, r_i] = -\left(\rho^2 + \sigma_\epsilon^2\right) - \frac{\rho\sigma}{\sqrt{n}}$, and $\mathrm{cov}[r_{i+j}, r_i] = 0$, $j > 1$. Therefore, from the covariogram we obtain $\sigma^2 = n(\mathrm{var}[r_{i+1}] + 2\mathrm{cov}[r_{i+1}, r_i])$ but we can only identify $\rho^2 + \sigma_\epsilon^2$ not the two quantities separately. There are just two equations in two unknowns and if ϵ_i is also Gaussian, then there is no more information. If there is a non-Gaussian distribution one can identify ρ using parametric restrictions. This is similar to the classical measurement error problem, Maddala, (1977, p 296).

our model, instead of n^{-1}, the appropriate scaling is $n^{\alpha-1}$. Such an estimator would converge to $\delta^2 + \int \omega^2(u)\,du$. Hence, this estimator would converge asymptotically to the integrated variance of the MS noise. Following Kristensen (2006), in the special case $\delta = 0$, we could also estimate $\omega(\cdot)$ at some fixed point τ using kernel smoothing,

$$\widehat{\omega}^2(\tau) = \frac{1}{2n^{1-\alpha}} \frac{\sum_{i=1}^{n} K_h(t_{i-1}-\tau)\left(\Delta Y_{t_{i-1}}\right)^2}{\sum_{i=1}^{n} K_h(t_{i-1}-\tau)(t_i - t_{i-1})}.$$

When the observations are equidistant, this simplifies to

$$\widehat{\omega}^2(\tau) = \sum_{i=1}^{n} K_h(t_{i-1}-\tau)\left(\Delta Y_{t_{i-1}}\right)^2 / 2n^{-\alpha}.$$

In the above, h is a bandwidth that tends to zero asymptotically and $K_h(.) = K(./h)/h$, where $K(.)$ is a kernel function satisfying some regularity conditions. If we also allow for endogeneity ($\delta \neq 0$), $\widehat{\omega}^2(\tau)$ estimates $\omega^2(\tau)$ plus a constant, and so we still see the pattern of diurnal variation. See Section 6 for implementation.

1.3 Estimation

We suppose that the parameter of interest is the quadratic variation of X on $[0,1]$, denoted $QV_X = \int_0^1 \sigma_t^2 dt$. Let

$$[Y,Y]^n = \sum_{i=1}^{n-1} \left(Y_{t_{i+1}} - Y_{t_i}\right)^2$$

be the realized variation (often called realized volatility) of Y, and introduce a modified version of it (*jittered* RV) as follows,

$$[Y,Y]^{\{n\}} = \frac{1}{2}\left(\sum_{i=1}^{n-K} \left(Y_{t_{i+1}} - Y_{t_i}\right)^2 + \sum_{i=K}^{n-1} \left(Y_{t_{i+1}} - Y_{t_i}\right)^2\right). \qquad (1.4)$$

This modification is useful for controlling the end effects that arise due to heteroscedasticity.

Our estimator of QV_X makes use of the same principles as the TSRV estimator in Zhang et al. (2005a). We split the original sample of size n into K subsamples, with the j^{th} subsample containing n_j observations. Introduce a constant β and c such that $K = cn^{\beta}$. The dependence of K on n is suppressed in the sequel. For consistency we will need $\beta > 1/2 - \alpha$. The optimal choice of β is discussed in the next section. By setting $\alpha = 0$, we get the condition for consistency in Zhang et al. (2005a), that $\beta > 1/2$.[5]

Let $[Y, Y]^{n_j}$ denote the j^{th} subsample estimator based on a K-spaced subsample of size n_j,

$$[Y, Y]^{n_j} = \sum_{i=1}^{n_j - 1} \left(Y_{t_{iK+j}} - Y_{t_{(i-1)K+j}}\right)^2, \; j = 1, \ldots, K,$$

and let

$$[Y, Y]^{avg} = \frac{1}{K} \sum_{j=1}^{K} [Y, Y]^{n_j}$$

be the averaged subsample estimator. To simplify the notation, we assume that n is divisible by K and hence the number of data points is the same across subsamples, $n_1 = n_2 = ... = n_K = n/K$. Let $\overline{n} = n/K$.

Define the adjusted TSRV estimator (*jittered* TSRV) as

$$\widehat{QV}_X = [Y, Y]^{avg} - \left(\frac{\overline{n}}{n}\right) [Y, Y]^{\{n\}}. \tag{1.5}$$

Compared to the TSRV estimator, this estimator does not involve any new parameters that would have to be chosen by the econometrician, so it is as easy to implement. The need to adjust the TSRV estimator arises from the fact that under our assumptions

[5]This condition is implicit in Zhang et al. (2005) in Theorem 1 (page 1400) where the rate of convergence is $\sqrt{K/\overline{n}} = c\sqrt{n^{2\beta-1}}$.

TSRV is not always consistent. The problem arises due to end-of-sample effects induced by heteroscedastic noise. For a simple example where the TSRV estimator is inconsistent, let us simplify the model to the framework of Zhang et al. (2005a), and introduce only heteroscedasticity in the noise, the exact form of which is to be chosen below. Let us evaluate the asymptotic bias of TSRV estimator.[6]

$$
\begin{aligned}
& n^{1/6} E\left\{\widehat{QV}_X^{TSRV} - QV_X\right\} \\
= \; & n^{1/6}\left\{E[u,u]^{avg} - \frac{\overline{n}}{n} E\left[u,u\right]^n\right\} + o\left(1\right) \\
= \; & c^{-1} n^{-1/2} \sum_{i=1}^{n-K} \left(\omega_{t_{i+K}}^2 \epsilon_{t_{i+K}}^2 + \omega_{t_i}^2 \epsilon_{t_i}^2\right) \\
& - \left(c^{-1} n^{-1/2} - n^{-5/6}\right) \sum_{i=1}^{n-1} \left(\omega_{t_{i+1}}^2 \epsilon_{t_{i+1}}^2 + \omega_{t_i}^2 \epsilon_{t_i}^2\right) + o\left(1\right) \\
= \; & n^{-5/6} \sum_{i=1}^{n-1} \left(\omega_{t_{i+1}}^2 \epsilon_{t_{i+1}}^2 + \omega_{t_i}^2 \epsilon_{t_i}^2\right) - c^{-1} n^{-1/2} \left\{\sum_{i=2}^{K} \omega_{t_i}^2 \epsilon_{t_i}^2 + \sum_{i=n-K+1}^{n-1} \omega_{t_i}^2 \epsilon_{t_i}^2\right\} + o\left(1\right).
\end{aligned}
$$

We see that the first and last K returns that are "ignored" by averaged subsampled realized volatility $[Y,Y]^{avg} \sim [u,u]^{avg}$ have to be off-set by a fraction of the noise of all returns, coming from $[Y,Y]^n \sim [u,u]^n$. For this bias correction to work, the volatility of the microstructure noise in the morning and afternoon has to be "close" to the volatility of the noise during the day. A simple counter-example that is motivated by our empirical section 6.3 is a parabola on $[0,1]$, $\omega^2\left(i/n\right) = a + \left(\frac{i}{n} - 0.5\right)^2/100$, where a is any constant. In this case simple calculations give that TSRV estimator is inconsistent,

$$
n^{1/6} \mathrm{E}(\widehat{QV}_X^{TSRV} - QV_X) = -\frac{1}{300} n^{1/6} + o\left(1\right).
$$

By contrast, *jittered* RV, $[Y,Y]^{\{n\}}$, mimics the structure of the volatility component

[6] For the reader to be able to follow our calculations in the next few lines, she should use the exact definition of $\overline{n}$, $\overline{n} = \frac{n-K+1}{K}$ that Zhang et al. (2005) use. For all other purposes differences between our and their definition are negligible.

that needs to be bias corrected for in $[Y,Y]^{avg}$, which is

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n-K}\left(\omega_{t_{i+K}}^2\epsilon_{t_{i+K}}^2+\omega_{t_i}^2\epsilon_{t_i}^2\right)$$

and so delivers a consistent estimator $\widehat{QV}_X$.

We remark that (1.5) is an additive bias correction and there is a nonzero probability that $\widehat{QV}_X < 0$. One can ensure positivity by replacing $\widehat{QV}_X$ by $\max\{\widehat{QV}_X, 0\}$, but this is not very satisfactory. Note, however, that we usually have $\widehat{QV}_X > \widehat{QV}_X^{TSRV}$ (except for when first and last subsamples have all flat prices and so $\widehat{QV}_X = \widehat{QV}_X^{TSRV}$), so the probability that $\widehat{QV}_X < 0$ is lower than the probability that $\widehat{QV}_X^{TSRV} < 0$.

1.4 Asymptotic Properties

The expansion for $[Y,Y]^{avg}$ and $[Y,Y]^n$ both contain terms due to the correlation between the measurement error and the latent returns. The main issues can be illustrated using the expansion of $[Y,Y]^{avg}$, conditional on the path of σ_t:

$$[Y,Y]^{avg} = \underbrace{QV_X}_{(a)} + \underbrace{2\frac{\delta\gamma_n}{K}\int_0^1 \sigma_t dt}_{(b)} + \underbrace{E\left[u,u\right]^{avg}}_{(c)} + O\left(\underbrace{\overline{n}^{-1/2}}_{(d)} + \underbrace{\sqrt{\frac{\overline{n}}{Kn^{2\alpha}}}}_{(e)}\right)Z, \qquad (1.6)$$

where $Z \sim N(0,1)$, while the terms in curly braces are as follows: (a) the probability limit of $[X,X]^{avg}$, which we aim to estimate; (b) the bias due to correlation between the latent returns and the measurement error; (c) the bias due to measurement error; (d) the variance due to discretization; (e) the variance due to measurement error.

Should we observe the latent price without measurement error, (a) and (d) would be the only terms. In this case, of course, it is better to use $[X,X]^n$, since that has an error of smaller order $n^{-1/2}$. In the presence of the measurement error, however, both $[Y,Y]^{avg}$ and $[Y,Y]^{n}$ are badly biased, the bias arising both from correlation between the latent returns and the measurement error, and from the variance of the measurement error. The largest term is (c), which satisfies

$$E\left[u,u\right]^{avg} = 2\overline{n}n^{-\alpha}\left(\int_0^1 \omega^2\left(u\right)du + \delta^2\right) + O\left(n^{-\alpha}+\overline{n}^{-1}\right) = O\left(\overline{n}n^{-\alpha}\right),$$

i.e., it is of order $\overline{n}n^{-\alpha}$. So without further modifications, this is what $[Y,Y]^{avg}$ would be estimating. Should we be able to correct that, the next term would be $2(\delta\gamma_n/K)\int \sigma_t dt$ arising from $E\left[X,u\right]^{avg}$. This second term is zero, however, if there is no correlation between the latent price and the MS noise, i.e., if $\delta = 0$. Interestingly when we use the TSRV estimator for bias correction of $E\left[u,u\right]^{avg}$, we also cancel this second term.

The asymptotic distribution of our estimator arises as a combination of two effects, measurement error and discretization effect. After correcting for the bias due to the measurement error (terms like b and c in eqn. 1.6), we still have the variation due to the measurement error (term e in eqn. 1.6). We can see that its contribution to the asymptotic distribution by observing how the estimator converges to the realized variance of the latent price X,

$$\sqrt{\frac{Kn^{2\alpha}}{\overline{n}}}\left(\widehat{QV}_X - [X,X]^{avg}\right) \Longrightarrow N\left(0, 8\delta^4 + 16\delta^2\int_0^1 \omega^2\left(u\right)du + 8\int_0^1 \omega^4\left(u\right)du\right), \tag{1.7}$$

The rate of convergence arises from $\mathrm{var}[u,u]^{avg} = O\left(\overline{n}/Kn^{2\alpha}\right)$. Both parts of the noise u, which are v and ε, contribute to the asymptotic variance. The first part of the asymptotic variance roughly arises from $\mathrm{var}[v,v]$, the second part from $\mathrm{var}[v,\varepsilon]$ (which is nonzero even though the correlation between both terms is zero), and the third part from $\mathrm{var}[\varepsilon,\varepsilon]$. If the measurement error is uncorrelated with the latent price, the first two terms disappear.

Should we observe the latent price without any error, we would still not know its quadratic variation due to observing the latent price only at discrete time intervals. This is another source of estimation error. From Theorem 3 in Zhang et al. (2005a) we have

$$\overline{n}^{1/2}\left([X,X]^{avg} - QV_X\right) \Longrightarrow MN\left(0, \frac{4}{3}\int_0^1 \sigma_t^4 dt\right), \tag{1.8}$$

where $MN(0,S)$ denotes a mixed normal distribution with conditional variance S independent of the underlying normal random variable.

The final result is a combination of the two results (1.7) and (1.8), as well as the fact that they are asymptotically independent. The fastest rate of convergence is achieved by choosing K so that the variance from the discretization is of the same order as the variance arising from the MS noise, so set $\overline{n}^{-1/2} = \sqrt{\overline{n}/Kn^{2\alpha}}$. The resulting optimal magnitude of K is such that $\beta = 2(1-\alpha)/3$. The rate of convergence with this rule is $\overline{n}^{-1/2} = n^{-1/6-\alpha/3}$. The slowest rate of convergence is $n^{-1/6}$, and it corresponds to large MS noise case, $\alpha = 0$. The fastest rate of convergence is $n^{-1/3}$, which corresponds to $\alpha = 1/2$ case. If we pick a larger β (and hence more subsamples K) than optimal, the rate of convergence in (1.7) increases, and the rate in (1.8) decreases and so dominates the final convergence result. In this case the final convergence is slower and only the first term due to discretization appears in the asymptotic variance (see (1.9)). Conversely, if we pick a smaller β (and hence K) than optimal, we get a slower rate of convergence and only the second term in the asymptotic variance ("measurement error" in (1.9)), which is due to the MS noise.

We obtain the asymptotic distribution of $\widehat{QV}_X$ in the following theorem

Theorem 1.4.1. *Suppose that* $\{X_t, t \in [0,1]\}$ *is a Brownian semimartingale satisfying (1.1). Suppose that* $\{\mu_t, t \in [0,1]\}$ *and* $\{\sigma_t, t \in [0,1]\}$ *are measurable and càdlàg* *processes, independent of the process $\{W_t, t \in [0,1]\}$. Suppose further that the observed price arises as in (1.2) with $\alpha \in [0, 1/2)$. Let the measurement error u_{t_i} be generated by (1.3), with ϵ_{t_i} i.i.d. mean zero and variance one and independent of the Gaussian process $\{W_t, t \in [0,1]\}$ with $E|\epsilon_{t_i}|^{4+\eta} < \infty$ for some $\eta > 0$. Then,*

$$V(\sigma)^{-1/2}\overline{n}^{1/2}\left(\widehat{QV}_X - QV_X\right) \Longrightarrow N(0,1),$$

$$V(\sigma) = \underbrace{\frac{4}{3}\int_0^1 \sigma_t^4 dt}_{discretization} + \underbrace{c^{-3}\left(8\delta^4 + 16\delta^2\int_0^1 \omega^2(u)\,du + 8\int_0^1 \omega^4(u)\,du\right)}_{measurement\ error} > 0 \text{ a.s.} \quad (1.9)$$

REMARKS.

1. The quantity $V(\sigma)$ collapses to the expression in Zhang et al. (2005a) when $\omega(.)$ is constant.

2. If one could find a consistent estimator $\widehat{V}(\sigma)$ such that $\widehat{V}(\sigma) - V(\sigma) = o(1)$ a.s., then the above theorem can be strengthened along the lines of Barndorff-Nielsen and Shephard to a feasible CLT, i.e., $\widehat{V}(\sigma)^{-1/2}\overline{n}^{1/2}(\widehat{QV}_X - QV_X) \Longrightarrow N(0,1)$ from which one could obtain confidence intervals for QV_X. Without assuming $\delta = 0$ or constant $\omega(.)$, the procedure of Zhang et al. (2005a), p. 1404, would work to estimate $V(\sigma)$.

3. The main statement of the Theorem 1.4.1 can also be written as

$$n^{1/6+\alpha/3}\left(\widehat{QV}_X - QV_X\right) \Longrightarrow MN(0, cV(\sigma)),$$

where $V(\sigma) = V_1(\sigma) + c^{-3}V_2$, with $V_1(\sigma)$ being the discretization error, while MN denotes a mixed normal distribution with conditional variance $cV(\sigma)$ independent of the underlying normal random variable. We can use this to find the value of c that would minimize the conditional asymptotic variance, $c_{opt}(\sigma) = (2V_2/V_1(\sigma))^{1/3}$, provided $V_1(\sigma) > 0$, resulting in the asymptotic conditional variance $(3/2^{2/3})V_2^{1/3}V_1^{2/3}(\sigma)$. If one has consistent estimators $\widehat{V}_j(\sigma) - V_j(\sigma) = o(1)$ a.s., $j = 1, 2$, then $\widehat{c}_{opt}(\sigma) = (2\widehat{V}_2(\sigma)/\widehat{V}_1(\sigma))^{1/3}$ is consistent in the sense that $\widehat{c}_{opt}(\sigma) - c_{opt}(\sigma) = o(1)$ a.s.

4. Suppose now that the measurement error is smaller than above and we have $\alpha \in [1/2, 1)$ instead of $\alpha \in [0, 1/2)$. Then, there is a consistency condition $\beta > 1/3$ that becomes binding and therefore optimal β allows the measurement error to converge faster than the discretization error. For $\beta = 1/3 + \Delta$ (where Δ small and positive) the rate of convergence is $\overline{n}^{-1/2} = n^{-(1-\beta)/2} = n^{-1/3+\Delta/2}$. Note that this is exactly the rate that occurs when there is no measurement error at all. So choose $\beta \in (1/3, 1)$. The conclusion of the Theorem 1.4.1 becomes

$$V(\sigma)^{-1/2} n^{(1-\beta)/2} \left(\widehat{QV}_X - QV_X\right) \Longrightarrow N(0, 1),$$

where $V_1(\sigma) = (4/3) \int \sigma_t^4 dt$. This can be shown by minor adjustments to the proofs.

5. What if $\alpha \geq 1$? This means that $[u, u]$ is of the same or smaller magnitude than $[X, X]$. In the case $\alpha = 1$ they are of the same order and identification breaks down. When $\alpha > 1$, realized volatility of observed prices is a consistent estimator of quadratic variation of latent prices, as measurement error is of smaller order. This is an artificial case and does not seem to appear in the real data.

How can we put this analysis in context? A useful benchmark for evaluation of the asymptotic properties of nonparametric estimators is the performance of parametric estimators. Gloter and Jacod (2001) allow for the dependence of the variance of i.i.d. Gaussian measurement error ρ_n on n and establish the Local Asymptotic Normality (LAN) property of the likelihood, which is a precondition to asymptotic optimality of the MLE. For the special case $\rho_n = \rho$ they obtain a convergence rate $n^{-1/4}$, thus allowing one to conclude that the MSRV and realized kernels can achieve the fastest possible rate. They also show that the rate of convergence is $n^{-1/2}$ if ρ_n goes to zero sufficiently fast, which is the rate when there is no measurement error at all. Our estimator has a rate $n^{-1/3+\Delta}$ when there is no measurement error, which is also the rate of convergence when the noise is sufficiently small. Also, Gloter and Jacod have that for "large" noise, the rate of convergence depends on the magnitude of the noise, similarly to our results. The rate of convergence and the threshold for the magnitude of the variance of the noise is different, though.

1.5 Simulation study

In this section we explore the behavior of the estimator (1.5) in finite samples. We simulate the Heston (1993) model:

$$\begin{aligned} dX_t &= (\mu_t - v_t/2)\,dt + \sigma_t dW_t \\ dv_t &= \kappa\,(\theta - v_t)\,dt + \gamma v_t^{1/2} dB_t, \end{aligned}$$

where $v_t = \sigma_t^2$, and W_t, B_t are independent standard Brownian motions.

For the benchmark model, we take the parameters of Zhang et al. (2005a): $\mu = 0.05$, $\kappa = 5$, $\theta = 0.04$, $\gamma = 0.5$. We set the length of the sample path to 23400 corresponding to the number of seconds in a business day, the time between observations corresponding to one second when a year is one unit, and the number of replications to be 100,000.[7] We set $\alpha = 0$. We choose the values of ω and δ so as to have a homoscedastic measurement error with variance equal to 0.0005^2 (again from Zhang et al. (2005a)), and correlation between the latent returns and the measurement error

[7] Note that in the theoretical part of the paper we had for brevity taken interval [0,1]. For the simulations we need the interval [0,1/250]. Suppose the parameter of interest is $\int_0^\tau \sigma_t^2 dt$, the quadratic variation of X on $[0,\tau]$. In that case the asymptotic conditional variance of the Theorem 1.4.1 becomes

$$V(\sigma) = \frac{4}{3}\tau \int_0^\tau \sigma_t^4 dt + c^{-3}\left(8\tau^2\delta^4 + 16\delta^2 \int_0^\tau \omega^2(u)\,du + 8\tau^{-1}\int_0^\tau \omega^4(u)\,du\right).$$

This follows by simple adjustments in the proofs. We take $\tau = 1/250$.

equal to -0.1. For this we use the identity

$$\operatorname{corr}(\Delta X_{t_i}, \Delta u_{t_i}) = \frac{E(\sigma)}{\sqrt{2E(\sigma^2)}} \frac{\delta}{\sqrt{\delta^2 + \omega^2}}$$

and the fact that for our volatility we have $E(\sigma) = \theta$, $\operatorname{var}(\sigma) = \theta\gamma^2/2\kappa$. We set $\beta = 2(1-\alpha)/3$. Figure 1.1 shows the common volatility path for all simulations.

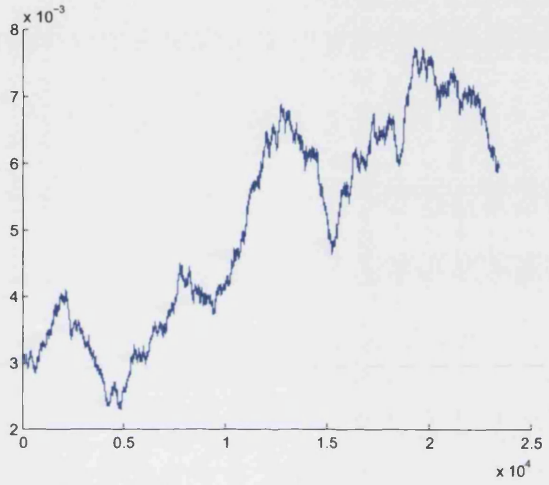


Figure 1.1: *The common volatility path for all simulations.*

First, we construct different models to see the effect of varying α and the number of observations within a day. We take the values of δ and ω that arise from the benchmark model, and then do simulations for the following combinations of α and n. When interpreting the results, we should also take into account that both of these parameters change the size of the variance of the measurement error. We measure the proximity of the finite sample distribution to the asymptotic distribution by the percentage errors of the interquartile range of $\overline{n}^{1/2}(\widehat{QV}_X - QV_X)$ compared to $1.3\sqrt{V}$, the value predicted by the distribution theory. We note that this is not the same as the MSE or variance of the estimator: it can be that a very efficient estimator can be poorly approximated by its limiting distribution and vice versa. This measure is easiest to interpret if we work with a fixed variance, i.e., when we condition on the volatility path. Hence, we simulate the volatility path for the largest number of observations, 23400, and perform all simulations using this one sample path of volatility. The last parameter to choose is K, the number of subsamples. This is the only parameter that an econometrician has to choose in practice. We examine four different values as follows (the expressions are all rounded to the closest integer):

$(2V_2/V_1)^{1/3}n^{\frac{2}{3}(1-\alpha)}$	asymptotically optimal rate and c	*Tables A.1 and A.2*
$n^{\frac{2}{3}(1-\alpha)}$	variation of above	*Tables A.3 and A.4*
$n^{\frac{2}{3}}$	variation of above	*Table A.5*
$\left(\frac{3RV^2}{2RQ}\right)^{1/3} n^{1/3}$	Bandi and Russell (2006a, eq. 24)	*Table A.6*

Table 1.1: Choices of K

Table A.1 contains the interquartile range errors (IQRs), in per cent, with the asymptotically optimal rate and constant (in terms of minimizing asymptotic mean squared error) for K. That is, we use $K = (2V_2/V_1)^{1/3}n^{2(1-\alpha)/3}$, rounded to the nearest integer, where V_1 and V_2 are discretization and measurement errors from (1.9). Table A.2 contains the values of K.

First of all, for small values of α, the percentage errors decrease with n as predicted by the theory. However, we do see some large errors, and from the values of K in Table A.2 we can guess this is due to the asymptotically optimal rule selecting very low c_{opt}. In fact, for the volatility path used here, $c_{opt} = (2V_2/V_1)^{1/3} = 0.0242$. Hence, another experiment we consider is an arbitrary choice $c = 1$. The next two tables (Table A.3 and A.4) contain the percentage errors and values of K that result from using $K = n^{2(1-\alpha)/3}$.

The performance of this choice is much better. We can see from Table A.3 that for small values of α, the asymptotic approximation improves with sample size. The sign of the error changes as α increases for given n, meaning that the actual IQR is below that predicted by the asymptotic distribution for small α and small n but this changes into the actual IQR being above the asymptotic prediction.

Another variant that does not include the unobservable α would be to use $K = n^{2/3}$.

Finally, we consider a method proposed by Bandi and Russell (2006a), which requires some discussion. They establish the exact mean squared error of TSRV under the assumptions of the independent additive noise model, and in addition they assume asymptotically constant volatility, i.e., $\int_{t_{i-1}}^{t_i} \sigma_u^2 du = \int_0^1 \sigma_u^2 du/n$ for each i, as well as $E(\epsilon^4) = 3E^2(\epsilon^2)$. Two assumptions are not satisfied in our simulation setup, the independence between the noise and the latent returns, as well as the assumption $\int_{t_{i-1}}^{t_i} \sigma_u^2 du = \int_0^1 \sigma_u^2 du/n$ for each i (see Figure 1.1). Therefore, this should be considered as another ad hoc selection method in our simulation setup. We note that this bandwidth choice results in an inconsistent estimator in our framework and in the framework of ZMA (2005a) (i.e., when $\alpha = 0$, $\beta > 1/2$ is required for consistency). Note that the choice K^{BR} was derived for $\widehat{QV}^{TSRV}$ without *jittering*, but this end-of-sample adjustment, though theoretically crucial, is negligible in simulations and, as we will see in the next section, also in real data. Table A.6 contains the IQR percentage errors and values of K that result from using $K^{BR} = (3RV^2/2RQ)^{1/3} n^{1/3}$, where RV is the realized variance, $RV = \sum(\Delta Y_{low})^2$ and RQ is the realized quarticity, $RQ = \frac{S}{3}\sum(\Delta Y_{low})^4$. Here, Y_{low} is low frequency (15 minute) returns, which gives $S = 24$ to be the number of low frequency observations during one day.

We see that the IQR errors of this choice get worse with sample size for small α, which reflects the inconsistency predicted by the theory. On the other hand the errors are small and improve with n for large α, i.e., when the noise is small. The performance is generally better than with asymptotically optimal K, except for cases that have both large n and small α, including the case $\alpha = 0$ usually considered in the literature. We notice that K^{BR} rule gives better results than the asymptotically optimal rule when it chooses a larger K, which is in most cases, but not all. In comparison to rules $K = n^{2(1-\alpha)/3}$ and $K = n^{2/3}$ (Tables A.3 and A.5, respectively), the performance of this choice is still disappointing, especially for small α. We conclude that in this setting the K^{BR} rule is not always the best choice according to our criterion.

It has been noted elsewhere that the asymptotic approximation can perform poorly, see Gonçalves and Meddahi (2005) and Aït-Sahalia, Zhang and Mykland (2005a).

From Tables A.1, A.3, and A.5 we see that magnitude of noise does not affect the quality of the asymptotic approximation. Although we see the interquartile range error having some relationship with α in Table A.3 and especially Table A.1, this is purely driven by changes in K. This is evidenced by Table A.5 where the rule for K does not depend on α and the respective error is close to constant for the same number of observations and different α. Another conclusion here is that a good rule for K does not necessarily have to depend on α, which is convenient for practical purposes.

In a second set of experiments we investigate the effect of varying ω, which controls the variance of the second part of the measurement error, for the largest sample size. Denoting by ω_b^2 the value of ω^2 in the benchmark model, we construct models with $\omega^2 = \omega_b^2, 4\omega_b^2, 8\omega_b^2, 10\omega_b^2$, and $20\omega_b^2$. The corresponding interquartile errors are 0.96%, 1.26%, 1.93%, 2.29%, and 4.64%.

In a third set of experiments we investigate the effect of varying δ, which controls the size of the correlation of the latent returns and measurement error. Denoting by δ_b^2 the value of δ^2 in the benchmark model, we construct models with δ^2 being from $0.01 \times \delta^2$ to $20 \times \delta^2$. The exact values of δ^2, as well as corresponding correlation between returns and increments of the noise, and the resulting interquartile errors are reported in Table A.7. We can see that when the number of observations is 23400, there is no strong effect from the correlation of the latent returns and measurement error on the approximation of the asymptotic interquartile range of the estimator.

1.6 Empirical analysis

To illustrate the above ideas, we perform a small empirical analysis. We discuss estimation of α, $\omega(.)$, and the quadratic variation of the latent price. The endogeneity parameter δ is unfortunately nonparametrically unidentified and so cannot be estimated. Its sole purpose is in allowing for flexible size and sign of endogeneity, with respect to which our estimator of quadratic variation is robust.

Figure A.1 in the appendix shows the volatility signature of the data we use, which is IBM transaction data, year 2005. The plot indicates that market microstructure noise is prevalent at the frequencies of $10-15$ minutes and higher. Since the volatility signature plot does not become negative, one cannot find evidence of endogeneity using the method of HL (2006). As pointed out already by HL(2006), this does not mean there is no endogeneity.

1.6.1 The Data

We use IBM transactions data for the whole year 2005. We employ the data cleaning procedure as in HL (2006), main paper and rejoinder. First, we use transactions from NYSE exchange only as this is the main exchange for IBM. Second, we use only transactions from 9:30AM to 4:00PM. Third, for transactions with the same time stamp, we use the average price. Fourth, we remove outliers as follows. If the price is too much above the ask price or too much below the bid, we remove it. Too high means more than spread above the ask, and too low means more than spread below the bid. Fifth, we remove days with less than 5 hours of trading (there were none). For discussion of the advantages of this procedure see HL (2006). The mean number of transactions per day in our cleaned data set is 4,484 (for comparison, there are 4,680 intervals of 5 seconds in the 6.5 hours between 9:30 and 16:00).

1.6.2 Estimation of α

The parameter that governs the magnitude of the microstructure noise, α, can be consistently estimated. Recall that the leading term of realized volatility $[Y,Y]^n$ is $[u,u]^n$ i.e.,

$$\begin{aligned}
[Y,Y]^n &= \sum_{i=1}^{n-1}(u_{t_{i+1}} - u_{t_i})^2 + o_p(n^{1-\alpha}) \\
&= n^{-\alpha}\sum_{i=1}^{n-1}(\omega_{t_{i+1}}\epsilon_{t_{i+1}} - \omega_{t_i}\epsilon_{t_i} + \delta\sqrt{n}(W_{t_{i+1}} - W_{t_i}))^2 + o_p(n^{1-\alpha}) \\
&= n^{1-\alpha}c + o_p(n^{1-\alpha})
\end{aligned}$$

for some positive constant c. It follows that

$$\log([Y,Y]^n/n) = -\alpha \log n + \log c + o_p(\log n).$$

We therefore estimate α by

$$\widehat{\alpha} = -\frac{\log\left([Y,Y]^n/n\right)}{\log(n)}, \tag{1.10}$$

see Linton and Kalnina (2007).

Although this is a consistent estimator for α, it has a bias that decays slowly. To reduce the bias, we estimate α over windows of 60 days instead of 1 day, i.e., we take our fixed interval $[0,1]$ to represent 3 months instead of 1 day. Figure 1.2 shows the estimates over the whole year 2005 where we roll the 60 day window by 1 day. We see that $\widehat{\alpha}$ varies between 0.64 and 0.7 with an average value of 0.67.

Although this is a consistent estimator for α, it is not precise enough to give a consistent estimator of n^{α}. As a consequence, this estimator cannot be used for consistent inference for $\widehat{QV}_x$. In Linton and Kalnina (2005) we provide a sharper bias adjusted version of $\widehat{\alpha}$, $\widehat{\alpha}^{adj}$, but the adjusted estimator is not feasible as it requires

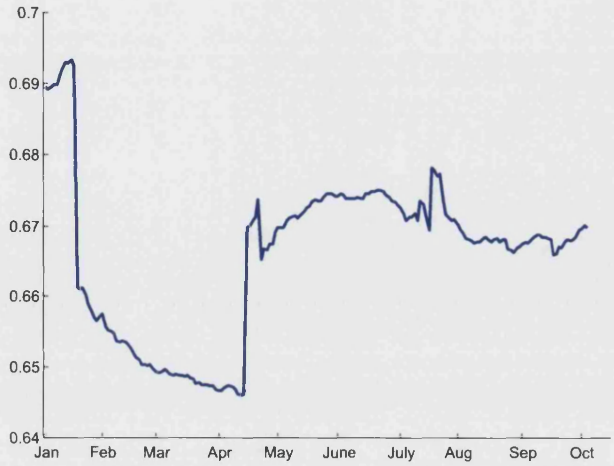


Figure 1.2: *Estimated α over a rolling window of 60 days (approx. 3 months). X axis shows the date of the first day in the window.*

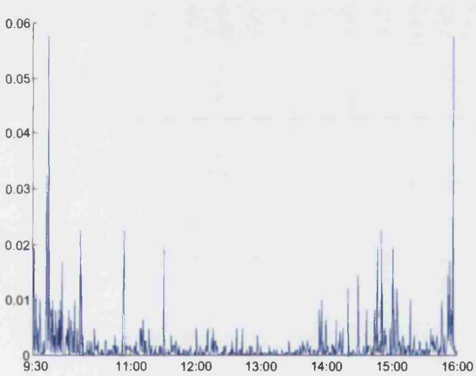
knowledge of $\omega(\tau)$. This last parameter can only be consistently estimated if $\alpha = 0$ and $\delta = 0$. The lack of precision in $\widehat{\alpha}$ also prevents us from developing a test of the null hypothesis $\alpha = 0$. Therefore, the deviations of $\widehat{\alpha}$ we see in Figure A.1 provide only a heuristic evidence that the true α is positive.

1.6.3 Estimation of Scedastic function $\omega(.)$

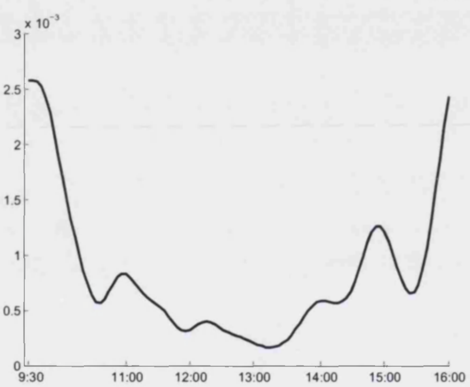
Now we estimate the function $\omega(\tau)$ that allows us to measure the diurnal variation of the MS noise. In the benchmark measurement error model this is a constant $\omega(\tau) \equiv \omega$ that can be estimated consistently by $\sum_{i=1}^{n-1}\left(Y_{t_{i+1}} - Y_{t_i}\right)^2/2n$ (Bandi and Russell (2006c), Barndorff-Nielsen et al. (2006), Zhang et al. (2005a)). In the special case $\alpha = 0$ and $\delta = 0$ this estimator would converge asymptotically to the integrated variance of the MS noise, $\int \omega^2(\tau)\,d\tau$. We can estimate the function $\omega^2(.)$ at a specific point τ using a simple generalization of the approach of Kristensen (2006) to the case with market microstructure noise. For equidistant observations, the estimator is

$$\widehat{\omega}^2(\tau) = \frac{\sum_{i=1}^{n} K_h\left(t_{i-1} - \tau\right)\left(\Delta Y_{t_{i-1}}\right)^2}{2n^{-\alpha}}. \tag{1.11}$$

We pick a random day, say 77th, which corresponds to 22nd of April. Assume $\alpha = 0$ and $\delta = 0$ and note that if these assumptions are not true, the level will be incorrect, while the diurnal variation will still be correct. Figure 1.3(b) shows the estimated function $\widehat{\omega}^2(\tau)$ using calendar time with 30 seconds frequency. We see that the variance of MS noise is far from being constant, and is closer to U-shape. Higher $\widehat{\omega}^2(\tau)$ at the beginning of the day and low values around 13:00 are displayed by virtually all days in 2005, while higher values of $\widehat{\omega}^2(\tau)$ at the end of the day are less common. Hence, overall, we confirm the findings of the empirical market microstructure literature that the intraday patterns are of U or reverse J shape (see references in the introduction).



(a) Squared returns



(b) Estimated function $\omega^2(.)$

Figure 1.3: *IBM transactions data, 22nd of April 2005.*

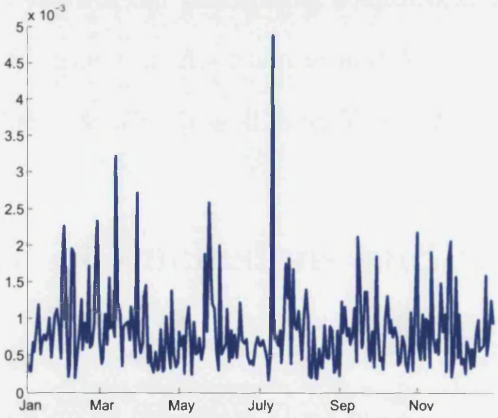
1.6.4 Estimation of Quadratic Variation

Our theory predicts that original *TSRV* estimator is asymptotically as good as our *jittered* version if intraday volatility pattern is "close enough" to constant volatility. Visual inspection of the estimated volatilities in the previous section suggest that there is some deviation from constant volatility, so one might call for adjustment to the *TSRV* estimator. How important is this adjustment in practice?

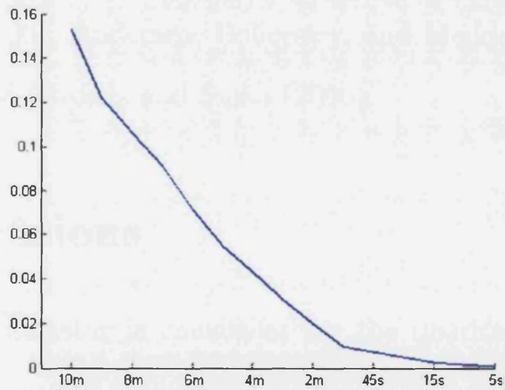
We check empirically the effect of *jittering* on daily point estimates of quadratic variation using IBM data in 2005. Figure 1.4(a) shows a plot of relative differences

$$\frac{\widehat{QV}_X - \widehat{QV}_X^{TSRV}}{\widehat{QV}_X^{TSRV}}$$

for every day in 2005 where we use tick time sampling (with 1-tick and $K = n^{2/3}$). The plot for 5 minute calendar time sampling (CTS) is very similar. The mean of these relative differences over all days is 0.0009. Figure 1.4(b) shows means of the relative difference for CTS, across different frequencies.[8] We see that, on average, for high frequencies, *jittering* makes very little difference. For lower frequencies the change is more visible. This arises from the fact that the *jittering* changes the $TSRV$ estimator on two subsamples only (see eqn. 1.12). The more subsamples there are, the less important our adjustment (this can also be achieved for any fixed frequency by using larger number of subsamples than our choice $K = n^{2/3}$).



(a) daily differences, 1-tick sampling



(b) average daily differences, CTS

Figure 1.4: *What is the relative difference* $\frac{\widehat{QV}_X - \widehat{QV}_X^{TSRV}}{\widehat{QV}_X^{TSRV}}$ *from our adjustment to the TSRV estimator?*

Another important observation is that *jittering* always increases the value of QV

[8]This average excludes October 27. On this day our estimator, when calculated on fequencies above 7 minutes, became several times bigger than TSRV estimator.

estimates, since we can write

$$\widehat{QV}_X^{TSRV} = \widehat{QV}_X + \frac{1}{2}\left(\sum_{i=1}^{K-1}\left(Y_{t_{i+1}} - Y_{t_i}\right)^2 + \sum_{i=n-K+1}^{n}\left(Y_{t_{i+1}} - Y_{t_i}\right)^2\right) > \widehat{QV}_X. \quad (1.12)$$

The more there is variation in the beginning of the day and the end of the day, the larger is the adjustment. This implies that *jittering* partly alleviates the problem that the usual TSRV estimator can sometimes become negative. With our data set, the only negative value (though very small) we saw was on February 28 when we calculated TSRV estimator with 10 minutes CTS frequency. The *jittered* version was positive.

We conclude that for most applications our estimator is very close to the $TSRV$ estimator, and so for practical applications plain $TSRV$ estimator can be used, without adjustment for heteroscedastic market microstructure noise. As a result, as far as point estimates are concerned, the existing empirical studies of $TSRV$ estimator are still valid in our theoretical framework. See, for example, investigations of forecasting performance in Aït-Sahalia and Mancini (2006), Andersen, Bollerslev, and Meddahi (2006), Bandi, Russell, and Yang (2007), and Ghysels and Sinko (2006).

1.7 Conclusions and Extensions

In this paper we showed that the TSRV estimator is consistent for the quadratic variation of the latent (log) price process when the measurement error is correlated with the latent price, although some adjustment is necessary when the measurement error is heteroscedastic. We also showed how the rate of convergence of the estimator depends on the magnitude of the measurement error.

Inference for TSRV estimator is robust to endogeneity of the measurement error. Provided the suggested adjustment to the estimator is implemented to preserve consistency, inference is also robust to heteroscedasticity of the noise. However, since the rate of convergence depends on the magnitude of the noise, inference is not robust to possible deviations from assumptions about this magnitude. We plan to investigate this question further.

Other examples where inference question needs to be solved include autocorrelation in measurement error (as in Aït-Sahalia, Mykland, and Zhang, 2006a), or other generalizations to the independent additive error model (Li and Mykland 2007). Gonçalves and Meddahi (2005) have recently proposed a bootstrap methodology for conducting inference under the assumption of no noise and shown that it has good small sample performance in their model. Zhang, Mykland, and Aït-Sahalia (2005b) have developed Edgeworth expansions for the TSRV estimator, and it would be very interesting to use this for analysis of inference using bootstrap. The results we have presented may be generalized to cover MSRV estimators and to allow for serial correlation in the error terms, although in both cases the notation becomes very complicated.

Chapter 2

Subsampling High Frequency Data

2.1 Introduction

This paper proposes the first automated method for conducting inference with high frequency data. In particular, it proposes to estimate the asymptotic variance of some estimator without relying on the exact expression of the asymptotic variance. In the traditional stationary time series framework, this task can be accomplished by bootstrap and subsampling variance estimators, but these are inconsistent with high frequency data.

A new subsampling method is developed, which enables to conduct inference for a general class of estimators that includes many estimators of integrated volatility. The question of inference on volatility estimates is important due to volatility being unobservable. For example, one might want to test whether volatility is the same on two different days, or in two different time periods within the same day. The latter corresponds to testing for diurnal variation in the volatility. Also, a common way of testing for jumps in prices is to compare two different volatility estimates, which converge to the same quantity under the null hypothesis of no jumps, but are different asymptotically under the alternative hypothesis of jumps in prices. Then, a consistent inferential method is needed to determine whether the two volatility estimates are significantly different.

To illustrate the robustness of the new method, this paper considers the example of inference problem for the integrated variance estimator of Aït-Sahalia et al. (2006a), in the presence of market microstructure noise. As several assumptions about the market microstructure noise are relaxed, the expression for the asymptotic variance becomes more complicated, and it becomes more challenging to estimate each component of the variance separately. On the other hand, the new subsampling method delivers consistent confidence intervals that are simple to calculate.

According to the fundamental theorem of asset pricing (see Delbaen and Schachermayer, 1994), the price process should follow a semimartingale. In this model, integrated variance (sometimes called integrated volatility) is a natural measure of variability of the price path (see, e.g. Andersen, Bollerslev, Diebold, and Labys, 2001). With moderate frequency data, say 5 or 15 minute data, this can be estimated by the so called Realized Variance (RV), a sum of squared returns. The nonparametric nature of Realized Variance and the simplicity of its calculation have made it popular among practitioners. It has been used for asset allocation (Fleming, Kirby, and Ostdiek, 2003), forecasting of Value at Risk (Giot and Laurent, 2004), evaluation of volatility forecasting models (Andersen and Bollerslev, 1998), and other purposes. The Chicago Board Options Exchange (CBOE) started trading S&P 500 Three-Month Realized Volatility options on October 21, 2008. Over the counter, these and other derivatives written on RV have been traded for several years. These financial products allow one to bet on the direction of the volatility, or to hedge against exposure to volatility. Pricing of these derivatives is done according to the theory of quadratic variation.

Suppose the log-price X_t follows a Brownian semimartingale process,

$$dX_t = \mu_t dt + \sigma_t dW_t, \tag{2.1}$$

where μ, σ, and W are the drift, volatility, and Brownian Motion processes, respectively. Our interest is in estimating volatility over some interval, say one day, which we normalize to be $[0, 1]$. The quantity of interest is captured by integrated variance, or quadratic variation over the interval, which is defined as

$$IV_X = \int_0^1 \sigma_s^2 ds.$$

Realized variance (or empirical quadratic variation) is a consistent estimator of integrated variance in infill asymptotics, i.e., when the the approximation is made as the time distance between adjacent observations shrinks to zero. According to this approximation, therefore, the estimation error in RV should be smaller for even higher frequency data than 5 minutes. Ironically, this is not the case in practice. For the highest frequencies, the data is more and more clearly affected by the bid-ask spread and other market microstructure frictions, rendering the semimartingale model inapplicable and RV inconsistent. Zhou (1996) proposed to model high frequency data as a Brownian semimartingale with an additive measurement error. This model can reconcile the main stylized facts of prices both in moderate and high frequencies. Zhang, Mykland, and Aït-Sahalia (2005) were the first to propose a consistent estimator of integrated variance in this model, in the presence of i.i.d. microstructure noise, which they named the Two Scale Realized Volatility estimator. Consistent estimators in this framework were also proposed by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008a), Christensen, Oomen, and Podolskij (2008), Christensen, Podolskij, and Vetter (2006), and Jacod, Li, Mykland, Podolskij, and Vetter (2007). Aït-Sahalia, Mykland, and Zhang (2006a) extend the Two Scale Realized Volatility estimator to the case of stationary autocorrelated microstructure noise, but do not propose an inference method. The problem with inference arises from the complicated structure of the asymptotic variance of the Two Scale Realized Volatility estimator. The method proposed in this paper can be used to conduct inference for the Two Scale Realized Volatility estimator in presence of not only autocorrelated, but also heteroscedastic measurement error. This allows the model to accommodate the stylized fact in the empirical market microstructure literature about the U-shape in observed returns and spreads.[1]

This new subsampling scheme is useful in practice when available estimators of the asymptotic variance are complicated and hence present difficulties in constructing confidence intervals. In such cases, a common procedure is to estimate the asymptotic variance as a sample variance of the bootstrap estimator. It turns out that even in the simple case of RV, this procedure is inconsistent, as the sample variance of the bootstrap estimator does not converge to the asymptotic variance of the original estimator, see Goncalvez and Meddahi (2008).

The subsampling method of Politis and Romano (1994) has been shown to be useful in many situations as a way of conducting inference under weak assumptions and without utilizing knowledge of limiting distributions. The basic intuition for constructing an estimator of the asymptotic variance is as follows. Imagine the standard setting of discrete time with long-span (also called increasing domain) asymptotics. Take some general estimator $\widehat{\theta}_n$ (think of i.i.d. $Y_i's$, a parameter of interest $\theta = E(Y)$, and $\widehat{\theta}_n = \frac{1}{n}\sum Y_i,$). Suppose we know its asymptotic distribution

$$\tau_n(\widehat{\theta}_n - \theta) \Longrightarrow N(0, V)$$

as $n \to \infty$, where $\Longrightarrow$ denotes convergence in distribution, and τ_n is the rate of

[1]See Andersen and Bollerslev (1997), Gerety and Mulherin (1994), Harris (1986), Lockwood and Linn (1990), and McInish and Wood (1992).

convergence when n observations are used. Suppose we would like to estimate V, in order to be able to construct confidence intervals for $\widehat{\theta}_n$. This can be done with the help of many subsamples, for which the estimator $\widehat{\theta}_n$ has the same asymptotic distribution. In particular, suppose we construct K different subsamples of $m = m(n)$ consecutive observations, starting at different values (whether they are overlapping or not is irrelevant here), where $m = m(n) \to \infty$ as $n \to \infty$ but $m/n \to 0$. Denote by $\widehat{\theta}_{n,m,l}$ the estimator $\widehat{\theta}_n$ calculated using the l^{th} block of m observations, with n being the total number of observations. Then, the asymptotic distribution of $\tau_m(\widehat{\theta}_{n,m,l} - \theta)$ is the same, i.e.

$$\tau_m \left(\widehat{\theta}_{n,m,l} - \theta\right) \Longrightarrow N(0, V) \tag{2.2}$$

for each subsample l, $l = 1, \ldots, K$. Hence, V can be estimated by the sample variance of $\tau_m \widehat{\theta}_{n,m,l}$ (with centering around $\widehat{\theta}_n$, a proxy for the true value θ). This yields the following estimator of V

$$\widehat{V} = \tau_m^2 \times \frac{1}{K} \sum_{l=1}^{K} \left(\widehat{\theta}_{n,m,l} - \widehat{\theta}_n\right)^2, \tag{2.3}$$

and we have

$$\widehat{V} \xrightarrow{p} V,$$

where $\xrightarrow{p}$ denotes convergence in probability. Notice that the estimator in (2.3) is like average of squared $\tau_m \left(\widehat{\theta}_{n,m,l} - \theta\right)$ over all subsamples, except that $\widehat{\theta}_n$ plays the role of θ. The difference between $\widehat{\theta}_n$ and θ is negligible because $\widehat{\theta}_n$ converges faster to θ than $\widehat{\theta}_{n,m,l}$ does.

It is shown that a direct application of the above method to the high frequency framework fails. This fact is illustrated for the RV example in model (2.1). That is, $\widehat{\theta}_n$ is taken to be Realized Variance and θ its probability limit, integrated variance. The intuition behind the failure is straightforward. The problem is that $\widehat{\theta}_{n,m,l}$ and $\widehat{\theta}_n$ do not converge to the same quantity and so (2.2) cannot be satisfied. The underlying reason is that the spot (or infinitesimal) volatility σ_t is changing over time. The estimator calculated on a small block cannot estimate the integrated variance θ, because θ contains information about spot volatility on the whole interval.

A novel subsampling scheme is proposed that can estimate the asymptotic variance of RV. Importantly, it can also be applied to the Two Scale Realized Volatility estimator of Aït-Sahalia et al. (2006a), in the presence of autocorrelated measurement error with diurnal heteroscedasticity. There are no alternative inferential methods available in the literature for this case. Moreover, this subsampling scheme can, under some conditions, estimate the asymptotic variance of a general class of estimators, which includes many estimators of the integrated variance.

The remainder of this paper is organized as follows. Section 2.2 describes the usual subsampling method of Politis and Romano (1994) and proposes a new subsampling method. It also introduces an alternative scheme that can be potentially useful. Section 2.3 shows how inference can be conducted for the Two Scale Realized Varinace estimator in the presence of autocorrelated and heteroscedastic microstructure noise. Section 2.4 applies the subsampling method to a general class of estimators. Section 2.5 investigates the numerical properties of the proposed method in a set of simulation experiments. Section 2.6 applies the method to high frequency stock returns. Section 2.7 concludes.

2.2 Description of Resampling Schemes

The aim of this section is to motivate and introduce a new subsampling scheme in a relatively simple framework. Since the proposed method does not change across models or estimators, the methodology and intuition behind it is illustrated with the example of Realized Volatility. The first subsection explains the failure of bootstrap and subsampling (Politis and Romano, 1994) methods for the estimation of the asymptotic variance of RV. The second subsection introduces a new subsampling scheme that can estimate the asymptotic variance of RV consistently. The third subsection describes an alternative scheme that is of theoretical interest, but which will not be used beyond the RV example.

We first describe the setting for the Realized Volatility example. Suppose that log-price X_t is the following Brownian semimartingale process

$$dX_t = \mu_t dt + \sigma_t dW_t, \tag{2.4}$$

where W_t is standard Brownian motion, the stochastic process μ_t is locally bounded, and σ_t is a càdlàg spot volatility process.[2] Suppose that we have observations on X on the interval $[0, T]$, where T is fixed. Without loss of generality set $T = 1$. Assume observation times are equidistant, so that the distance between observations is $1/n$. The asymptotic scheme is infill as $n \to \infty$.

Suppose the quantity of interest is integrated variance (also called integrated volatility),

$$IV_X = \int_0^1 \sigma_s^2 ds. \tag{2.5}$$

IV_X is a random variable depending on the realization of the volatility path $\{\sigma_t, t \in [0, 1]\}$. The usual estimator of IV_X is the Realized Variance (often called Realized Volatility)

$$RV_n = \sum_{i=1}^{n} \left(X_{i/n} - X_{(i-1)/n}\right)^2. \tag{2.6}$$

This satisfies

$$\sqrt{n}\,(RV_n - IV_X) \Longrightarrow MN(0, V) \tag{2.7}$$

$$V = 2IQ = 2\int_0^1 \sigma_s^4 ds$$

where $MN(0, V)$ denotes a mixed normal distribution with random conditional vari-

[2]In other words, the sample paths of the volatility process are left continuous with right limits.

ance V independent of the underlying normal distribution.[3] The convergence (2.7) follows from Barndorff-Nielsen and Shephard (2002) and Jacod (2006), and is stable in law, see Aldous and Eagleson (1978). Stable convergence is slightly stronger than the usual convergence in distribution. Stable asymptotics are particularly convenient because it permits division of both sides of (2.7) by the square root of any consistent estimator of V to obtain standardized asymptotic distribution for conducting inference on RV_n.

In fact, for the Realized Variance example, inference can be conducted relatively easily. Barndorff-Nielsen and Shephard (2002) propose to estimate V as twice the realized quarticity, $\widetilde{V} = 2IQ_n$, where realized quarticity is sum of fourth powers of returns, properly scaled,

$$IQ_n = \frac{n}{3}\sum_{i=1}^{n}\left(X_{i/n} - X_{(i-1)/n}\right)^4. \tag{2.8}$$

The estimator $\widetilde{V}$ is consistent for V in the sense that $\widetilde{V}/V \xrightarrow{p} 1$. This result allows the construction of consistent confidence intervals for QV_X. For example, a two-sided level α interval is given by $\widetilde{C}_\alpha = RV_n \pm z_{\alpha/2}\widetilde{V}^{1/2}/\sqrt{n}$, where z_α is the α quantile from a standard normal distribution, and this has the property that $\Pr[IV_X \in \widetilde{C}_\alpha] \to 1-\alpha$. Mykland and Zhang (2006,7) have proposed an alternative estimator of V that is more efficient than $\widetilde{V}$ under the sampling scheme (2.4) and can also be used to construct intervals based on the studentized limit theory.

The next subsection explains why the usual bootstrap and subsampling methods cannot be used to estimate V in this framework. Then, a new subsampling method is introduced. Section 2.2 concludes with description of an alternative subsampling scheme.

[3] In other words, the limiting p.d.f. is of the form $f(x) = \int \phi_{0,v}(x) f_V(v)dv$, where f_V denotes the p.d.f. of V and $\phi_{0,v}(x) = \exp(-x^2/2v^2)/\sqrt{2\pi v}$

2.2.1 Failure of the Traditional Resampling Schemes

Recently, Goncalvez and Meddahi (2008) have proposed a bootstrap algorithm for RV. They use the i.i.d. and wild bootstrap applied to studentized RV. They show that resampling the studentized RV gives confidence intervals for RV with better properties than the $2IQ_n$ estimator of asymptotic variance. An essential feature of their procedures is reliance on an estimator of the asymptotic variance, which is not always available. A more widely used bootstrap procedure is to estimate asymptotic variance as the sample variance of the bootstrap statistic. This procedure is simple, but inconsistent in the high frequency framework, as even in the simple case of RV, the bootstrap estimator has a different asymptotic variance than the original estimator, see Goncalvez and Meddahi (2008). This means that confidence intervals constructed using the usual bootstrap method are inconsistent.

We now consider the popular method of Politis and Romano (1994). This subsampling scheme fails in our setting, highlighting the difference that high frequency framework brings. It is however instructive to consider, as subsequent methods proposed use the same underlying idea.

Let $\widehat{\theta}_n$ be the RV calculated on the full sample, and let $\widehat{\theta}_{n.m.l}$ be the RV calculated on the l^{th} block of m observations,[4]

$$\widehat{\theta}_{n.m.l} = \sum_{i=m(l-1)}^{ml} \left(X_{i/n} - X_{(i-1)/n}\right)^2,$$

see Figure 2.1. In the above, $0 < l \leq K$, where K is the number of subsamples, $K = \lfloor n/m \rfloor$.

Assumption 5.3.1 of Politis, Romano and Wolf (1999) is satisfied, i.e., the sampling distribution of $\tau_n(\widehat{\theta}_n - \theta)$ converges weakly. Therefore, in the setting of stationary

[4]For simplicity, all subsampling schemes in this paper are presented with non-overlapping subsamples. However, it is inconvenient to display non-overlapping subsamples in Figures, so Figures 2.1, 2.2, and 2.3 show maximum overlap versions of the subsampling schemes.

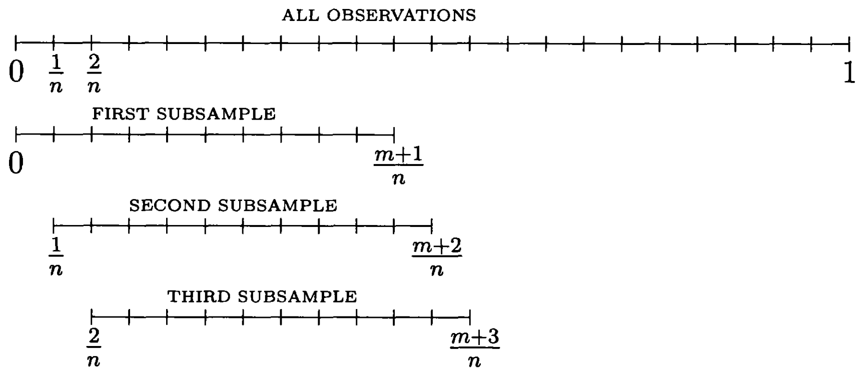


Figure 2.1: *The Subsampling Scheme of Politis and Romano (1994)*

and mixing processes, V should be approximated well by

$$\widehat{V}_{PR} = m \times \frac{1}{K} \sum_{l=1}^{K} \left(\widehat{\theta}_{n,m,l} - \widehat{\theta}_n \right)^2 .$$

However, in our setting, it is easy to see that $\widehat{V}_{PR}$ does not converge to V. The estimator on the full sample converges to the true value, $\widehat{\theta}_n \to^p \theta$. On the other hand, the estimator on a subsample converges to zero. This is because each high frequency return is of order $n^{-1/2}$, so a sum of m squared returns is of order $m/n \to 0$. Thus, $\left(\widehat{\theta}_{n,m,l} - \widehat{\theta}_n \right)^2$ converges to θ and $\widehat{V}_{PR}$ is asymptotically equal to $m\theta^2$. Notice that the value θ^2 is not related to V, which is the parameter of interest. A formal proof of the following proposition is in the appendix,

Proposition 2.2.1. *Let X satisfy (2.4) and $\widehat{\theta}_n$ be the Realized Variance defined in (2.6). Let $m \to \infty$ and $m/n \to 0$ as $n \to \infty$. Then,*

$$\widehat{V}_{PR} - m\theta^2 = o_p(m) .$$

A crucial ingredient of the subsampling method of Politis and Romano is that $\widehat{\theta}_{n,m,l}$ and $\widehat{\theta}_n$ estimate the same quantity. A direct application of their method to high frequency framework violates this basic principle. Part of the reason is the different rates of magnitude. This could be accounted for by using $\frac{m}{n}\widehat{\theta}_{n,m,l}$ instead of $\widehat{\theta}_{n,m,l}$. In this case, it still holds that $\frac{n}{m}\widehat{\theta}_{n,m,l} - \widehat{\theta}_n \overset{p}{\nrightarrow} 0$. This is because $\frac{m}{n}\widehat{\theta}_{n,m,l}$ estimates the spot variance $\sigma^2(\cdot)$ at some point, instead of the integrated variance θ.[5] Therefore, the underlying reason for the failure of the subsampling method of Politis and Romano is the fact that the spot variance changes over time.

2.2.2 The New Subsampling Scheme

We now introduce and motivate the new subsampling scheme. The current subsection describes this scheme for the RV example, and Section 2.3 applies it to the Two Scale Realized Volatility estimator. Section 2.4 applies this subsampling scheme to a more general class of estimators.

In the subsampling scheme of Politis and Romano (1994), the problem was that the estimator on a subsample $\widehat{\theta}_{n,m,l}$ was centered at "the wrong quantity". In the formula

$$\widehat{V}_{PR} = m \times \frac{1}{K}\sum_{l=1}^{K}\left(\widehat{\theta}_{n,m,l} - \widehat{\theta}_n\right)^2,$$

the quantity $\widehat{\theta}_n$ plays the role of θ, but the problem is that the leading term in $\widehat{\theta}_{n,m,l}$ is integrated variance over a shrinking interval,

$$\theta_l = \int_{(l-1)m/n}^{lm/n} \sigma^2(u)\, du. \tag{2.9}$$

Thus, $\widehat{\theta}_{n,m,l}$ either converges to zero or the spot volatility depending on whether it is scaled by n/m, but in any case it cannot estimate θ, the integrated volatility over the whole interval $[0,1]$. Therefore, $\widehat{\theta}_{n,m,l} - \widehat{\theta}_n$ does not converge to zero, causing $\widehat{V}_{PR}$ to explode.

[5]For estimation of the spot variance using RV on a shrinking interval, see Foster and Nelson (1996), Andreou and Ghysels (2002), Mikosch and Starica (2005), and Kristensen (2008).

Consider an alternative approach. We aim to center estimators at θ_l, in order to extract the information about the variance of $\widehat{\theta}_{n,m,l}$. The leading term of the variance of $\widehat{\theta}_{n,m,l}$ is

$$V_l = 2 \int_{(l-1)m/n}^{lm/n} \sigma_u^4 du.$$

It is of course not equal to V, which we want to estimate, but we can use the fact that these add up to V over subsamples,

$$V = 2\int_0^1 \sigma_u^4 du = \sum_{l=1}^{K} V_l.$$

Given the additive structure of $\widehat{V}$, this approach can still give a consistent estimator of V, despite volatility changing over time. The only question left is, how to obtain an estimator of the centering factor θ_l. So consider using two subsamples, one with length J and one with length m, such that J is of smaller order than m. Then, both $\frac{n}{m}\widehat{\theta}_{n,m,l}$ and $\frac{n}{J}\widehat{\theta}_{n,J,l}$ estimate the spot variance, but they have different convergence rates. This in turn means one can be used to center the other. To simplify the presentation, we use notation $\widehat{\theta}_l^{long}$ and $\widehat{\theta}_l^{short}$ instead of $\widehat{\theta}_{n,m,l}$ and $\widehat{\theta}_{n,J,l}$.

Since the rate of convergence of $\frac{n}{J}\widehat{\theta}^{short}$ is $\sqrt{J}$, the estimator of V becomes

$$\widehat{V}_{sub} = J \times \frac{1}{K}\sum_{l=1}^{K}\left(\frac{n}{J}\widehat{\theta}_l^{short} - \frac{n}{m}\widehat{\theta}_l^{long}\right)^2 \tag{2.10}$$

where $K = \lfloor n/m \rfloor$. $\widehat{\theta}_l^{short}$ and $\widehat{\theta}_l^{long}$ are Realized Variances calculated on the short subsample with J observations, and the long subsample with m observations. Figure 2.2 provides a graphical illustration. The corresponding time intervals used are $\left[\frac{(l-1)m}{n}, \frac{(l-1)m+J}{n}\right]$ and $\left[\frac{(l-1)m}{n}, \frac{lm}{n}\right]$, so the expressions for estimators on subsamples

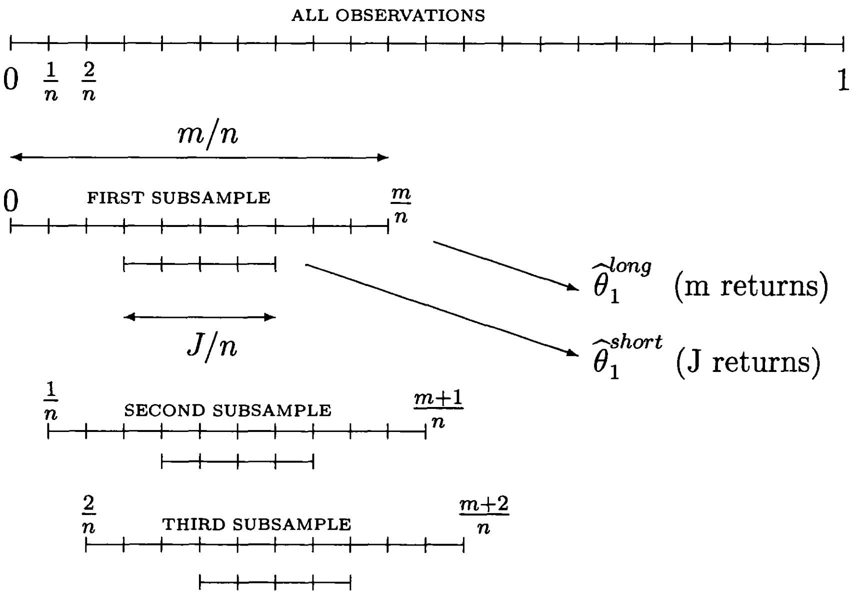


Figure 2.2: *The New Subsampling Scheme*

become

$$\begin{aligned}\widehat{\theta}_l^{short} &= \sum_{i=1}^{J}\left(X_{\frac{(l-1)m+i}{n}} - X_{\frac{(l-1)m+i-1}{n}}\right)^2 \\ \widehat{\theta}_l^{long} &= \sum_{i=1}^{m}\left(X_{\frac{(l-1)m+i}{n}} - X_{\frac{(l-1)m+i-1}{n}}\right)^2.\end{aligned}$$

For an arbitrary volatility process, $nJ^{-1}\widehat{\theta}_l^{short}$ and $nm^{-1}\widehat{\theta}_l^{long}$ cannot be guaranteed to be close. For example, if the volatility process has a large jump on the interval covered by $\widehat{\theta}_l^{long}$, but not covered by $\widehat{\theta}_l^{short}$, then $nJ^{-1}\widehat{\theta}_l^{short}$ and $nm^{-1}\widehat{\theta}_l^{long}$ can differ substantially. Therefore, some kind of smoothness condition on the volatility paths is needed. Importantly, we do not require differentiable sample paths. It can be shown that a sufficient condition is to assume that volatility itself evolves like a Brownian semimartingale. This is a common way of modeling volatility in practice.

Assumption A1. *The volatility process $\{\sigma_t, t \in [0, 1]\}$ is a Brownian semimartingale* *of the form*

$$d\sigma_t = \widetilde{\mu}_t dt + \widetilde{\sigma}_t d\widetilde{W}_t$$

where $\widetilde{W}_t$ *is standard Brownian motion, the stochastic process* $\widetilde{\mu}_t$ *is locally bounded and the stochastic process* $\widetilde{\sigma}_t$ *is càdlàg.*

Proposition 2.2.2. *Suppose (A1) holds and X satisfies (2.4). Let* $\widehat{\theta}_n$ *be the Realized Variance defined in (2.6),* $m \to \infty$, $J \to \infty$, $m/n \to 0$, $J/m \to 0$, *and* $J^2/n \to 0$ *as* $n \to \infty$. *Then,*

$$\widehat{V}_{sub} \xrightarrow{p} V.$$

Sections 3 and 4 show that Proposition 2.2.2 can be extended to more general settings than RV in a Brownian semimartingale model. This is because the subsampling method does not rely on the exact form of V, which it estimates.

2.2.3 An Alternative Subsampling Scheme

This subsection presents an alternative subsampling scheme that is of theoretical interest. In general, it can be applied to cases when the asymptotic variance of an estimator on a sub-block of lower frequency observations has the same structure as the asymptotic variance of the estimator on the full sample. This scheme will not be used in further sections due to its inability to estimate the asymptotic variance of the Two Scale estimator in the presence of autocorrelated noise. We now describe it for the RV example.

Consider the following subsampling scheme. On every block of m observations, calculate the estimator $\widehat{\theta}^n$ twice as follows. First, calculate it using all m observations, denote it as $\widehat{\theta}_l^{fast}$. Then, calculate the estimator $\widehat{\theta}^n$ using every Q^{th} price observation in the block of m observations, and denote it as $\widehat{\theta}_l^{slow}$.

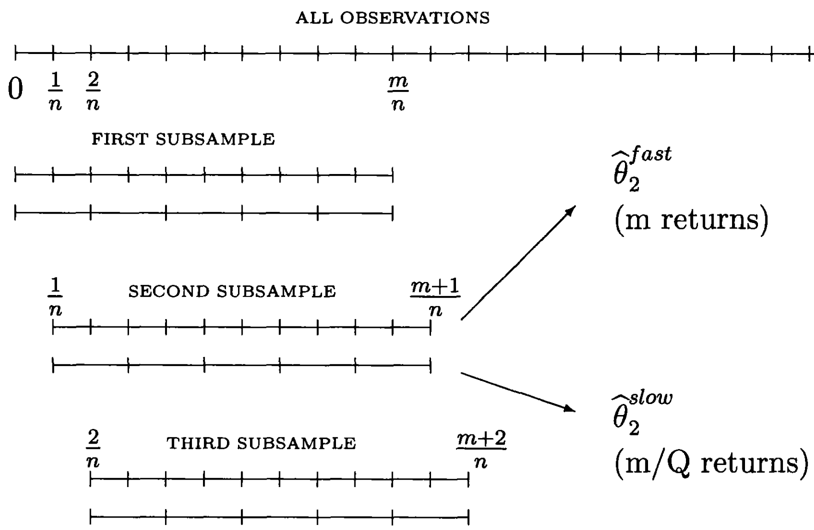


Figure 2.3: *An Alternative Subsampling Scheme.*

The corresponding expressions for RV calculated on these subsamples are

$$\begin{aligned}\widehat{\theta}_k^{fast} &= \sum_{i=1}^{m}\left(X_{\frac{i+m(k-1)}{n}} - X_{\frac{i-1+m(k-1)}{n}}\right)^2\\ \widehat{\theta}_k^{slow} &= \sum_{i=1}^{\lfloor m/Q\rfloor}\left(X_{\frac{iQ+m(k-1)}{n}} - X_{\frac{(i-1)Q+m(k-1)}{n}}\right)^2.\end{aligned}$$

Now, $\widehat{\theta}_l^{fast}$ can be used to center the $\widehat{\theta}_l^{slow}$, because they both converge to (2.9), and because $\widehat{\theta}_l^{fast}$ converges to (2.9) faster than $\widehat{\theta}_l^{slow}$ does. See Figure 2.3 for a graphical illustration. The estimator of V becomes

$$\widehat{V}_a = \frac{m}{Q}\times\frac{1}{K}\sum_{l=1}^{K}\left(\frac{n}{m}\widehat{\theta}_l^{slow} - \frac{n}{m}\widehat{\theta}_l^{fast}\right)^2 = \frac{n}{Q}\sum_{l=1}^{n/m}\left(\widehat{\theta}_l^{slow} - \widehat{\theta}_l^{fast}\right)^2$$

This construction shows that lower frequency data can be used to achieve the same effect as taking a shorter block of observations. In our RV example, sparse observations still convey all the features of the model, so this subsampling scheme delivers a consistent estimator of V.

Proposition 2.2.3. *Suppose X satisfies (2.4). Let $m \to \infty$, $Q \to \infty$, $m/n \to 0$, and*

$Q/m \to 0$ as $n \to \infty$. Then,

$$\widehat{V}_a \xrightarrow{p} V. \tag{2.11}$$

Remarks. *1. Brownian semimartingale model (2.4) assumes the paths of X are continuous. Instead, suppose now that X is a Brownian semimartingale with jumps. In other words, define X on some probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\right)$ as the process $dX_t = \mu_t dt + \sigma_t dW_t + dJ_t$. The continuous part is as in (2.4) and J_t is some jumps process, see for example Aït-Sahalia and Jacod (2008). In that case, the asymptotic variance of RV contains jumps, and the subsampling estimator $\widehat{V}_a$ only estimates consistently the continuous part of the V. In particular, suppose $m \to \infty$, $Q \to \infty$, $m/n \to 0$, and $Q/m \to 0$ as $n \to \infty$. Then,*

$$\widehat{V}_a \xrightarrow{p} 2\int_0^1 \sigma_u^4 du + 4 \sum_{p:T_p \in [0,1]} Y_{T_p}^2 \left(\sqrt{\kappa_p} U_p \sigma_{T_{p-}} + \sqrt{1-\kappa_p} U_p' \sigma_{T_p}\right)^2$$

where $\kappa_p, p = 1, 2, ...$ are uniform random variables, independent from F; U_p, $U_p', p = 1, 2, ...$ are standard normal random variables independent from F and from κ_p, $p = 1, 2, ...$; T_p, $p = 1, 2, ...$ are jump times. This shows the inconsistency of $\widehat{V}_a$ because the random variables U_p and U_p' do not appear in the expression of the asymptotic variance of RV. If X and σ do not jump together, $\widehat{V}_a$ is unbiased, conditionally on F because the random variable $\left(U_p + U_p'\right)^2$ has expectation one. This illustrates the fact that the subsampling method needs V to be continuous in time. This prevents a situation when there is some feature of V that is only represented by one subsample. See also discussion of Assumption A6(ii) in Section 2.4.

2. Suppose, as in Remark 1, that X is a Brownian semimartingale with jumps. In that case, integrated volatility can be estimated by Bipower Variation (see Barndorff-Nielsen and Shephard, 2007),

$$\widehat{\theta}^n = n^{\frac{r+l}{2}-1} \sum_{i=1}^{n} \left|X_{\frac{i}{n}} - X_{\frac{i-1}{n}}\right|^r \left|X_{\frac{i-1}{n}} - X_{\frac{i-2}{n}}\right|^l.$$

Then, the subsampling estimator of the asymptotic variance V of $\widehat{\theta}^n$ is only consistent if V does not contain jumps. This happens if $\max(r, l) < 1$.

The subsampling scheme is similar in structure to the one in Lahiri, Kaiser, Cressie, and Hsu (1999). They similarly use two grids for subsampling to predict stochastic cumulative distribution functions in a spatial framework. However, they assume that the underlying process is stationary and their asymptotic framework is mixed infill and increasing domain.

This alternative subsampling method can also be applied to noisy diffusion setting, as long as noise is independent in time. However, if noise is autocorrelated and this autocorrelation appears in the expression of V, this subsampling scheme will not be able to estimate V. This is due to sparsely sampled data not containing all the needed information about autocorrelations. Therefore, $\widehat{\theta}_l^{slow}$ is not be able to replicate that part in V, which pertains to autocorrelations.

2.3 Inference for the Two Scale Realized Volatility Estimator

This section shows how the new subsampling scheme can be applied to the Two Scale Realized Volatility estimator of integrated variance proposed by Aït-Sahalia et al. (2006a). Although only this example is discussed in detail, this subsampling scheme could also be applied to other integrated variance estimators in the presence of market microstructure noise, such as Multiscale estimators of Zhang et al. (2005) and Aït-Sahalia et al. (2006a), Realized Kernels of Barndorff-Nielsen et al. (2008a), and pre-averaging estimator of Jacod et al. (2007).

Stock price data at highest frequencies is well known to be affected by market microstructure noise. For example, trades are not executed in practice at the efficient price. Typically, they are executed either at the prevailing bid or ask price. Therefore, observed transaction prices alternate between bid and ask prices (the so-called bid-ask bounce), creating negative autocorrelation in observed returns, which is a stylized fact in high frequency data. This was the motivation for Zhou (1996) to introduce an additive market microstructure noise model where the observed log-price Y is a sum of a Brownian semimartingale component X and an i.i.d. noise ϵ, see (2.13) below. In this model, observed log-returns display negative first order autocovariance,

$$
\begin{aligned}
& \operatorname{Cov}\left(\Delta Y_{i/n}, \Delta Y_{(i-1)/n}\right) \qquad (2.12)\\
= \; & \operatorname{Cov}\left(\Delta X_{i/n} + \epsilon_{i/n} - \epsilon_{(i-1)/n}, \Delta X_{(i-1)/n} + \epsilon_{(i-1)/n} - \epsilon_{(i-2)/n}\right)\\
= \; & -\operatorname{Var}\left(\epsilon_{(i-1)/n}\right).
\end{aligned}
$$

Another stylized fact is that Realized Variances calculated at the highest frequencies become very large. This is in contradiction to the Brownian semimartingale model, where RV has roughly the same expectation irrespective of the frequency, at which it is calculated. Also, RV should converge to IV_X when higher and higher frequencies are used. This difficulty lies behind the underlying reason for the common practice not to calculate Realized Variance at higher frequencies than 5 or 15 minutes. The problem with this approach is that it implies discarding most of the available data. There are only 72 five minute returns in a day, and only 24 fifteen minute returns in a day, while the available high frequency data is usually measured in thousands. In order to be able to use all the available data, one has to work with a model that can accommodate the above stylized facts.

Zhang, Mykland, and Aït-Sahalia (2005) were the first to introduce a consistent estimator of integrated variance of the efficient price IV_X within the additive measurement error model of Zhou (1996). Their model is as follows. Log-price X is a continuous Brownian semimartingale (2.4). Observations are contaminated by some additive measurement error, so there are discrete observations on the noisy log-price

Y available where

$$Y_t = X_t + \epsilon_t. \tag{2.13}$$

The noise ϵ_t is *i.i.d.*, zero mean with variance $\mathrm{Var}(\epsilon) = \omega^2$ and $\mathrm{E}\epsilon^4 < \infty$, and independent from the latent log-price X_t. In this model, Zhang et al. (2005) propose the following consistent estimator for the integrated variance of X_t,

$$\widehat{\theta}_n = [Y,Y]^{(G_1)} - \frac{\overline{n}_{G_1}}{n} [Y,Y]^{(1)}, \tag{2.14}$$

where, for any parameter b,

$$\begin{aligned} [Y,Y]^{(b)} &= \frac{1}{b} \sum_{i=1}^{n-b} \left(Y_{(i+b)/n} - Y_{i/n}\right)^2 \\ \overline{n}_b &= \frac{n-b+1}{b}. \end{aligned}$$

Notice that $[Y,Y]^{(1)}$ coincides with the RV estimator, while $[Y,Y]^{(G_1)}$ consists of lower frequency returns. In particular, $[Y,Y]^{(G_1)}$ consists of returns calculated from prices that are G_1 high frequency observations apart. Thus, time distance is n^{-1} between high frequency observations and $G_1 n^{-1}$ between lower frequency observations. In empirical applications, a common choice for G_1 is such that the lower frequency returns are sampled at 5 minutes. Zhang et al. (2005) call the above estimator the Two Scale Realized Volatility (TSRV) estimator. They derive the following asymptotic distribution of the estimator,

$$n^{1/6}\left(\widehat{\theta}_n - \theta\right) \Rightarrow \sqrt{V} Z$$

where the asymptotic (conditional) variance takes the form

$$V = \underbrace{c\frac{4}{3}\int_0^1 \sigma_u^4 du}_{signal} + \underbrace{8c^{-2}\omega^4}_{noise}, \tag{2.15}$$

i.e., it consists of a signal part, which is due to the efficient price, and a noise part. In the above, Z is a standard normal random variable, independent from V, and c is the constant in $G_1 = \lfloor cn^{2/3} \rfloor$. With i.i.d. noise, V can be estimated component by component. $\text{Var}(\epsilon) = \omega^2$ can be estimated using the following estimator proposed by Bandi and Russell (2008),

$$\widehat{\omega^2} = \frac{RV}{2n} \xrightarrow{p} \omega^2.$$

We saw in Section 2.2 that in a model without noise, integrated quarticity $\int \sigma_u^4 du$ can be estimated by realized quarticity defined in (2.8). This becomes more difficult in the presence of noise. However, Barndorff-Nielsen et al. (2008a) have proposed an estimator for $\int \sigma_u^4 du$, which is consistent in presence of i.i.d. noise, see Section 2.5.

This model is for i.i.d. noise, so the noise is assumed to be homoscedastic. A well known stylized fact in empirical market microstructure literature is that intradaily spreads (difference between bid and ask price) and intradaily stock price volatility are described typically by a U-shape (See Footnote 2 for references). In other words, prices are more volatile in mornings and afternoons than at noon; spreads are also larger in mornings and afternoons. Figure 2.4(a) presents an estimate of heteroscedasticity function $\omega^2(\cdot)$ for transaction prices of Microsoft stock, averaged over all days in year 2006. The diurnal variation is evident.

Kalnina and Linton (2006) introduce diurnal heteroscedasticity in the microstructure noise in model (2.13). Suppose the efficient log-price X is the same as above in (2.13), but the noise displays unconditional heteroscedasticity. In particular, suppose

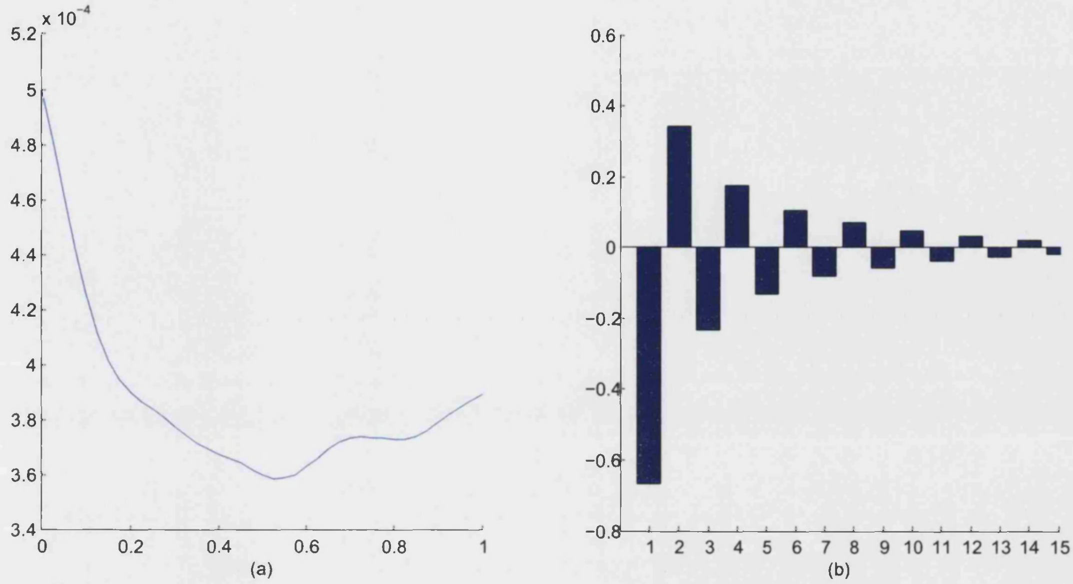


Figure 2.4: *Properties of returns of Microsoft (MSFT) stock. Returns are constructed from transaction prices over the whole year 2006. See Section 2.6 for data cleaning procedures. Panel (a) shows the estimated heteroscedasticity function* $\omega(\cdot)$ *, averaged over all days in 2006. Panel (b) shows the autocorrelogram of returns calculated in tick time.*

the noise ϵ_t satisfies

$$\epsilon_t = \omega(t)\, u_t \tag{2.16}$$

where $\omega(t)$ is a nonstochastic differentiable function of time t, and u_t is i.i.d. with $\mathrm{E}(u_t) = 0$, and $\mathrm{Var}(u_t) = 1$. As a result of this generalization, the asymptotic variance of $\widehat{\theta}_n$ changes to

$$V = \underbrace{c\frac{4}{3}\int_0^1 \sigma_u^4 du}_{signal} + \underbrace{8c^{-2}\int_0^1 \omega^4(u)\, du.}_{noise}$$

In this model, the previous estimator of the noise part of V ceases to be consistent as

$$\widehat{\omega^2} = \frac{RV}{2n} \xrightarrow{p} \int_0^1 \omega^2(u)\, du,$$

so, by Jensen's inequality, its square would be always strictly smaller than the target

$\int \omega^4 (u)\, du$ as long as there is any diurnal variation at all. Kalnina and Linton (2006) show that $\omega(\cdot)$ can be estimated at any fixed point τ using kernel smoothing,

$$\widetilde{\omega}^2(\tau) = \frac{1}{2}\sum_{i=1}^{n} K_h(t_{i-1} - \tau)\left(\Delta Y_{t_{i-1}}\right)^2.$$

In the above, h is a bandwidth that tends to zero asymptotically and $K_h(.) = K(./h)/h$, where $K(.)$ is a kernel function satisfying some regularity conditions. This suggests estimating the noise part of V by

$$8c^{-2}\int_0^1 \widetilde{\omega}^4(u)\, du.$$

As we saw earlier in (2.12), the i.i.d. measurement error model is consistent with negative first order autocorrelations in the observed returns. However, returns can sometimes exhibit autocorrelation beyond the first lag in practice. For example, Figure 2.4(b) graphs the autocorrelogram of the returns of the Microsoft stock for the whole year 2006. We see that Microsoft stock returns display strong negative autocorrelation well beyond the first lag. While the model (2.13) does generate a negative first autocorrelation, it implies that any further autocorrelations have to be zero. Since increments of a Brownian semimartingale are uncorrelated in time, any such autocorrelation has to be due to noise ϵ_t.[6]

Aït-Sahalia et al. (2006a) generalize the i.i.d. measurement error model (2.13) in a different direction. They allow for autocorrelated stationary microstructure noise. In particular, they make the following assumption about the noise.

Assumption A2. *The noise ϵ_t is independent from the efficient log-price X_t. Also, when viewed as a process in index i, ϵ_{t_i} is stationary and strong mixing with the mixing coefficients decaying exponentially*

[6] In a Brownian semimartingale model, the only source of autocorrelations of increments is drift, which is negligible for high frequencies.

In model (2.13) with ϵ_t satisfying Assumption A2, Aït-Sahalia et al. (2006a) propose the following consistent estimator for the integrated variance of X_t,

$$\widehat{\theta}_n = [Y,Y]^{(G_1)} - \frac{\overline{n}_{G_1}}{\overline{n}_{G_2}} [Y,Y]^{(G_2)} \tag{2.17}$$

where G_1 and G_2 satisfy the following assumption,

Assumption A3. *The G_1 parameter of the Two Scale Realized Volatility estimator $\widehat{\theta}_n$ defined by (2.17) satisfies $G_1 = \lfloor cn^{2/3} \rfloor$ for some constant c. G_2 parameter is such that* $\text{Cov}\left(\epsilon_0, \epsilon_{G_2/n}\right) = o\left(n^{-1/2}\right)$, $G_2 \to \infty$, $G_2/G_1 \to 0$.[7]

The Two Scale Realized Volatility estimator defined by (2.17) is a more general than the one in (2.14), which is a special case when $G_2 = 1$ and $G_1 \to \infty$ as $n \to \infty$. Aït-Sahalia et al. (2006a) show that the new TSRV estimator $\widehat{\theta}_n$ has the same asymptotic properties except it has a more complicated asymptotic variance,

$$V = \underbrace{c\frac{4}{3}\int_0^1 \sigma_u^4 du}_{signal} + \underbrace{8c^{-2}\text{Var}(\epsilon)^2 + 16c^{-2} \lim_{n\to\infty} \sum_{i=1}^{n} \text{Cov}\left(\epsilon_0, \epsilon_{i/n}\right)^2}_{noise}, \tag{2.18}$$

where c is the constant in $G_1 = \lfloor cn^{2/3} \rfloor$.

The literature does not provide any estimator for V or an alternative method for constructing confidence intervals for $\widehat{\theta}_n$. Here we can estimate the asymptotic variance of the Two Scale Realized Volatility estimator $\widehat{\theta}_n$ using the subsampling scheme.

Theorem 2.3.1. *Suppose model (2.13) holds, and ϵ_{t_i} satisfy Assumption A2. Let $\widehat{\theta}_n$ be the TSRV estimator defined by (2.17), with parameters G_1 and G_2 that satisfy Assumption A3. Let V be defined by (2.18). Assume $\{\sigma\}$ and $\{\mu\}$ are independent*

[7]The restriction on $\text{Cov}\left(\epsilon_0, \epsilon_{G_2/n}\right)$ should be considered in the light of the fact that Assumption A2 implies that there exists a constant ϕ such that, for all i,

$$\left|\text{Cov}\left(\epsilon_{i/n}, \epsilon_{(i+l)/n}\right)\right| \leq \phi^l \text{Var}(\epsilon).$$

of $\{W\}$. *Let* $J \to \infty$, $m \to \infty$, $J/m \to 0$, $m/n \to 0$, $G_1/J \to 0$ *and* $Jmn^{-5/3} \to 0$ *as* $n \to \infty$. *Then,*

$$\widehat{V}_{sub} \xrightarrow{p} V$$

where

$$\widehat{V}_{sub} = Jn^{-2/3} \times \frac{1}{K}\sum_{l=1}^{K}\left(\frac{n}{J}\widehat{\theta}_l^{short} - \frac{n}{m}\widehat{\theta}_l^{long}\right)^2 \tag{2.19}$$

with $K = \lfloor n/m \rfloor$.

In above, $\widehat{\theta}_l^{short}$ is simply $\widehat{\theta}^n$ calculated on a smaller block of J observations inside the l^{th} larger block of m observations, with exactly the same parameters G_1 and G_2 as $\widehat{\theta}^n$ uses. See Figure 2.2 for an illustration. In particular,

$$\widehat{\theta}_l^{short} = [Y,Y]_l^{(G_1)} - \frac{\overline{J}_{G_1}}{\overline{J}_{G_2}}[Y,Y]_l^{(G_2)}$$

where

$$\begin{aligned} [Y,Y]_l^{(G_i)} &= \frac{1}{G_i}\sum_{i=1}^{J-G_i}\left(Y_{(l-1)m/n+(i+G_i)/n} - Y_{(l-1)m/n+i/n}\right)^2, \; i=1,2 \\ \overline{J}_{G_i} &= \frac{J-G_i+1}{G_i}, \quad i=1,2. \end{aligned}$$

One obtains $\widehat{\theta}_l^{short}$ by substituting J for m above. In Figure 2.2, the version with maximum overlap is presented. In practice, it is much quicker to compute the no overlap version, for which Theorem 2.3.1 is formulated. While this does not alter the conclusion of Theorem 2.3.1, the maximum overlap version is slightly more efficient. In this case, $\widehat{V}_{sub}$ is defined by (2.19) with $K = n - m + 1$.

To the author's knowledge, this is the only available method in the literature to construct confidence intervals for the Two Scale Realized Volatility estimator when the noise is autocorrelated. Similarly, one can apply this method to Multi-Scale estimator of Aït-Sahalia et al. (2006a) when microstructure noise is autocorrelated. The advantage of using Multi-Scale estimator is that it has the optimal rate of convergence $n^{1/4}$.

However, the above model of Aït-Sahalia et al. (2006a) rules out any diurnal heteroscedasticity of the noise. When both autocorrelation and heteroscedasticity is taken into account, we have

Lemma 2.3.2. *Suppose the observed price satisfies $Y_{i/n} = X_{i/n} + \epsilon_{i/n}$ where the efficient log-price X_t follows a Brownian semimartingale process (2.4) and microstructure noise $\epsilon_{i/n}$ satisfies*

$$\epsilon_t = \omega(t)\, u_t$$

where $\omega(\cdot)$ is a differentiable, nonstochastic function of time, u_t satisfies Assumption A2 and $\mathrm{Var}(u_t) = 1$. Then, $\widehat{\theta}_n$ defined in (2.17) satisfies

$$n^{1/6}\left(\widehat{\theta}_n - \theta\right) \Rightarrow \sqrt{V} Z$$

where

$$V = c\frac{4}{3}\int_0^1 \sigma_u^4 du + 8c^{-2}\int_0^1 \omega^4(u)\, du + 16c^{-2}\int_0^1 \omega^4(u)\, du \lim_{n\to\infty}\sum_{i=1}^{n} \mathrm{Cov}\left(\epsilon_0, \epsilon_{i/n}\right)^2.$$

In this case of autocorrelated and heteroscedastic noise, Theorem 2.3.1 easily generalizes and subsampling again delivers consistent estimate of V. This is because both are special cases of the consistency result of the subsampling estimator in the general case, which is described in the next section. To estimate this more complicated V, exactly the same formula $\widehat{V}_{sub}$ should be used as for homoscedastic case. In this model, this is the only available method in the literature to construct confidence intervals for the Two Scale Realized Volatility estimator.

Importantly, this section illustrates robustness of the subsampling estimator of V

across different sets of assumptions. Moreover, it is also easy to implement. All that is necessary is to compute $\widehat{\theta}_n$ on several sub-blocks of observations. It seems that the subsampling estimator $\widehat{V}_{sub}$ would be consistent for V under even more general assumptions than considered above, for example, in the case when autocorrelations of the noise are changing through time.

2.4 Inference for a General Estimator

This section shows how to use the new subsampling scheme (as described in Sections 2.2.2 and 2.3) to conduct inference for a general class of estimators of volatility measures. A set of assumptions is introduced and explained, under which subsampling delivers a consistent estimate of the asymptotic variance of an estimator $\widehat{\theta}_n$. As we shall see, there are two essential ingredients for subsampling method to work. One is additivity over subsamples of the asymptotic variance of $\widehat{\theta}_n$. The second is that the asymptotic distribution of $\widehat{\theta}_n$ calculated on a block of observations is similar, in a sense explained below, to the asymptotic distribution of $\widehat{\theta}_n$ calculated using all available data.

We do not assume a specific process for X. It could be a pure diffusion or a diffusion contaminated with noise, as long as the regularity assumptions below are satisfied. All arguments in this section are made conditional on the volatility path $\{\sigma_u, u \in [0,1]\}$. Suppose there is an estimator $\widehat{\theta}_n$, for which the asymptotic distribution is known to be as follows

$$\tau_n\left(\widehat{\theta}_n - \theta\right) \Rightarrow \sqrt{V}Z. \tag{2.20}$$

In the above, τ_n is a known rate of convergence of $\widehat{\theta}_n$. For example, $\tau_n = n^{1/2}$ for RV, $\tau_n = n^{1/6}$ for the TSRV estimator. Z is a random variable that is known to satisfy $\mathrm{E}(Z) = 0$ and $\mathrm{Var}(Z) = 1$. A consistent estimator of V thus enables a researcher to construct consistent confidence intervals for $\widehat{\theta}_n$.

We recall the subsampling scheme introduced in Section 2.2.2. Divide the total number of returns into blocks of m consecutive returns. Thus, we obtain $\lfloor n/m \rfloor$ subsamples. Denote by $\widehat{\theta}_l^{long}$ the estimator $\widehat{\theta}_n$ calculated using all m returns of the l^{th} block, $l = 1, ..., \lfloor n/m \rfloor$. Denote by $\widehat{\theta}_l^{short}$ the estimator $\widehat{\theta}_n$ calculated using only J returns of the l^{th} block, where $J < m$. See Figure 2.2 in Section 2.2.2 for a graphical illustration.

In order to guarantee that $\frac{n}{J}\widehat{\theta}_l^{short}$ and $\frac{n}{m}\widehat{\theta}_l^{long}$ converge to the same quantity, despite being defined on different time intervals, we need to impose some smoothness on the volatility paths. In particular, we use the following assumption.

Assumption A4. *(2.20) holds, where θ and V are the following functions of the volatility path $\{\sigma_u, u \in [0,1]\}$,*

$$\begin{aligned} \theta &= \int_0^1 g_1(\sigma(u))\, du \\ V &= \int_0^1 g_2(\sigma(u))\, du \end{aligned}$$

where g_1,$g_2 \in C^1[0,1]$ and σ is a Brownian semimartingale as in (2.4).

For example, we obtain integrated variance IV_X with $g_1(u) = \sigma^2(u)$ and the asymptotic variance of RV with $g_2(\sigma(u)) = 2\sigma^2(u)$.

The type of estimators that are likely to satisfy assumptions of this section are those that are *approximately* additive over subsamples, i.e.,

$$\widehat{\theta}_n = \sum_{l=1}^{\lfloor n/m \rfloor} \frac{m}{J}\widehat{\theta}_l^{short} + o_p(1) \tag{2.21}$$

or

$$\widehat{\theta}_n = \sum_{l=1}^{\lfloor n/m \rfloor} \widehat{\theta}_l^{long} + o_p(1). \tag{2.22}$$

All currently available estimators of integrated variance and related quantities satisfy this additivity property. We also impose the following assumption, which ensures that estimators on subsamples are mixing.

Assumption A5. *For any fixed n, the returns process* $\left\{R_{i/n}^{(n)}\right\}_{i=1,\dots,n}$ *with* $R_{i/n}^{(n)} = X_{i/n} - X_{(i-1)/n}$ *is strong mixing. Also,* $\widehat{\theta}_n = \phi\left(R_{1/n}^{(n)}, R_{2/n}^{(n)}, \dots, R_1^{(n)}\right)$ *where* $\phi : \mathbb{R}^n \longmapsto \mathbb{R}$.

For example, suppose X is a Brownian semimartingale as in (2.4). Then, Assumption A5 is satisfied if we assume $\{\mu\}, \{\sigma\} \perp \{W\}$ and consider all arguments conditional on $\{\mu_t, \sigma_t, t \in [0,1]\}$. This means that the conclusion about consistency of $\widehat{V}_{sub}$ of Theorem 2.4.1 below holds, conditionally on $\{\mu_t, \sigma_t, t \in [0,1]\}$. Hence, the same conclusion also holds unconditionally.

As discussed in previous sections, $\widehat{\theta}_l^{long}$ and $\widehat{\theta}_l^{short}$ do not estimate θ, since they use only information about the volatility path on a small time interval, whereas the volatility is changing throughout the interval $[0,1]$. Let us denote by θ_l^{long} and θ_l^{short} the respective quantities they estimate, and by V_l^{short} and V_l^{long} what can be thought of as their asymptotic variances. They can be defined as follows,

$$\begin{aligned} \theta_l^{short} &= \int_{(l-1)m/n}^{[(l-1)m+J]/n} g_1(\sigma(u))\, du, \quad V_l^{short} = \int_{(l-1)m/n}^{[(l-1)m+J]/n} g_2(\sigma(u))\, du \\ \theta_l^{long} &= \int_{(l-1)m/n}^{lm/n} g_1(\sigma(u))\, du, \qquad V_l^{short} = \int_{(l-1)m/n}^{[(l-1)m+J]/n} g_2(\sigma(u))\, du. \end{aligned} \tag{2.23}$$

Finally, we make the following assumption,

Assumption A6. *For every n, define* θ_l^{short} *and* V_l^{short} *by (2.23), and define a triangular array*

$$\zeta_l^{(n)} = \frac{n}{J}\left[\tau_n^2\left(\widehat{\theta}_l^{short} - \theta_l^{short}\right)^2 - V_l^{short}\right].$$

The array $\left\{\zeta_j^{(n)}\right\}$ *satisfies the following conditions*

(i)

$$\text{as } n \to \infty, \ \sup_l \mathrm{E}\left(\zeta_l^{(n)}\right) \to 0,$$

(ii) $\left\{\zeta_j^{(n)}\right\}$ *is* L^p *bounded for some* $p > 1$.

We now discuss Assumption (A6). The appendix contains verification of Assumption (A6) for the TSRV estimator, as this is how Theorem 2.3.1 is proved.

Assumption A6 (i) can be written equivalently as follows,

$$\text{as } n \to \infty, \ \sup_l \mathrm{E}\left(\left(V_l^{short}\right)^{-1} \tau_n^2 \left(\widehat{\theta}_l^{short} - \theta_l^{short}\right)^2\right) \to 1,$$

as long as V_l^{short} is of order J/n. In other words, assumption A6 (i) requires that the square of the standardized statistic $\widehat{\theta}_l^{short}$ has asymptotic expectation one. On the full sample, we know from (2.20) that standardized $\widehat{\theta}^n$ is asymptotically a random variable Z with $\mathrm{E}\left(Z^2\right) = 1$. Therefore, a sufficient condition for Assumption A6 (i) to hold is that the asymptotic distribution of $\widehat{\theta}_l^{short}$ satisfies the same condition on a subsample. Roughly speaking, we need the estimator on a subsample, $\widehat{\theta}_l^{short}$, to behave similarly to the estimator on a full sample, $\widehat{\theta}_n$.

Assumption A6 (ii) is a stronger assumption, and it illustrates the main idea of the subsampling method. Recall the basic idea of subsampling as described in the introduction of the paper. Roughly speaking, in a stationary world, the way subsampling estimates V is by constructing many random variables with V as their asymptotic variance. In our nonstationary case, continuity in time plays the role of stationarity as it ensures that the same feature in V is estimated by many subsamples. Assumption A6 (ii) effectively imposes V_j^{short} to be of order J/n, i.e., that there is enough continuity in V with respect to time. Apart from this consideration, assumption A6 (ii) requires existence of moments. This is not an issue for a Brownian semimartingale model due to the local boundedness assumption on the drift and volatility functions, but becomes a constraint if X also contains other components. For example, consider a model where observations are sampled from a Brownian semimartingale with an additive noise ϵ. In this model, corresponding moments have to be assumed on ϵ for assumption A6 (ii) to hold. In the case of the TSRV estimator discussed below, $L^{4+\varepsilon}$ boundedness of ϵ is needed, which is exactly what has been assumed by the authors of TSRV estimator to derive its asymptotic distribution.

We have the following result.

Theorem 2.4.1. *Assume (A4), (A5), and (A6). Then,*

$$\widehat{V}_{sub} \xrightarrow{p} V$$

where

$$\widehat{V}_{sub} = \frac{Jm}{n^2} \sum_{l=1}^{\lfloor n/m \rfloor} \tau_n^2 \left(\frac{n}{J} \widehat{\theta}_l^{short} - \frac{n}{m} \widehat{\theta}_l^{long} \right)^2 .$$

Importantly, exactly the same formula is applied to all models and estimators, which satisfy the above assumptions. All that is necessary to calculate the estimator for V is to calculate the estimator $\widehat{\theta}^n$ on several subsamples, as well as to know the convergence rate τ_n. In particular, $\widehat{V}_{sub}$ simplifies to formula for the RV in (2.10) with $\tau_n = \sqrt{n}$, and to the formula for the Two Scale Realized Volatility estimator in (2.19) with $\tau_n = n^{1/6}$.

2.5 Simulation Study

In this section numerical properties of the proposed estimator are studied for the example of TSRV estimator of Aït-Sahalia et al. (2006a) in the case of i.i.d. or autocorrelated microstructure noise.

The observed price Y_t is a sum of the efficient log-price X_t and microstructure noise u_t. The paths of the efficient log-price are simulated from the Heston (1993) model:

$$
\begin{aligned}
dX_t &= (\alpha_1 - v_t/2)\, dt + \sigma_t dW_t \\
dv_t &= \alpha_2 (\alpha_3 - v_t)\, dt + \alpha_4 v_t^{1/2} dB_t
\end{aligned}
$$

where $v_t = \sigma_t^2$, W_t and B_t are independent Brownian Motions. The parameters of the efficient log-price process X are chosen to be the same as in Zhang et al. (2005). They are $\alpha_1 = 0.05$, $\alpha_2 = 5$, $\alpha_3 = 0.04$, and $\alpha_4 = 0.5$ (unit of time is one year). We simulate 35,000 observations over one week, i.e., five business days of 6.5 hours each. This is motivated by the fact that GE stock has on average 35,000 observations per week in year 2006, see Section 2.6.

The microstructure noise is simulated as an MA(1) process

$$
u_{\frac{i}{n}} = \epsilon_{\frac{i}{n}} + \rho \epsilon_{\frac{i-1}{n}}, \ \epsilon \sim N\left(0, \frac{\omega^2}{1+\rho^2}\right).
$$

Four different values of ρ are considered, $\rho = 0$, -0.3, -0.5, and $\rho = -0.7$.

Size of the noise is an important parameter. Denote $\mathrm{Var}(u) = \omega^2$. Denote the noise-to-signal ratio by

$$
\xi^2 = \frac{\omega^2}{\sqrt{\int_0^1 \sigma_u^4 du}}.
$$

Results are simulated for two different noise-to-signal ratios, which are suggested by Barndorff-Nielsen et al. (2008a) and are $\xi^2 = 0.001$ and 0.0001. These are motivated by the careful empirical study of Hansen and Lunde (2006), who investigate 30 stocks of Dow Jones Industrial Average. The volatility path is fixed over simulations to facilitate comparisons. The volatility path used is plotted in Figure 5. Varying volatility path across simulations does not affect the theory nor simulation results.

The parameters of the TSRV estimator and of the subsampling procedure are

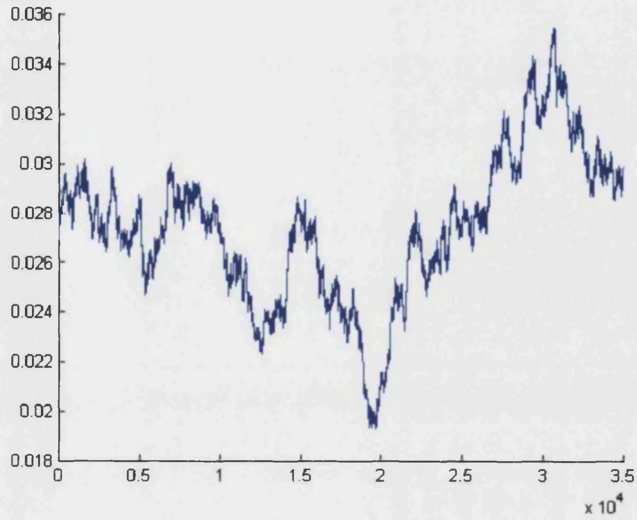


Figure 2.5: *Simulated volatility sample path.*

chosen as follows. We set $G_1 = 100$, which in our data corresponds to 5 minute lower frequency. This is a very popular choice in practice. We set $G_2 = 10$ in all simulations. Two values of J are considered. First is $J = 2G_1 = 200$, second is $J = 5G_1 = 500$. For m parameter, three different values are considered, $m = 4J, 10J$, and $15J$.

The literature does not propose ways of estimating asymptotic variance of TSRV when noise is autocorrelated or diurnal. However, in the case of i.i.d. noise, there is an alternative, and this will serve as a benchmark for the simulation results. In the case of i.i.d noise, the expression for asymptotic variance V of TSRV estimator is

$$V = c\frac{4}{3}\int_0^1 \sigma_u^4 du + 8c^{-2}\left[\mathrm{Var}\left(u\right)\right]^2$$

and the alternative is to estimate each component of V separately. The easiest component to estimate is $\left[\mathrm{Var}\left(u\right)\right]^2$. A popular estimator of $\mathrm{Var}\left(u\right) = \omega^2$ is

$$\widehat{\omega^2} = \frac{RV}{2n}.$$

This has been proposed by, for example, Bandi and Russell (2006, 2008). To estimate integrated quarticity IQ (the first term in V) in the presence of microstructure noise is more difficult. A consistent estimator in the presence of i.i.d. noise has been proposed by Barndorff-Nielsen et al. (2008a)

$$\widehat{IQ}_{BNHLS}\left(\widetilde{\delta}, S\right) = \max\left[\left(\widehat{\theta}_n^*\right)^2, \frac{1}{\widetilde{n}}\sum_{j=1}^{\widetilde{n}} \widetilde{\delta}^{-2}\left(y_{j,\cdot}^2 - 2\widetilde{\omega^2}\right)\left(y_{j-2,\cdot}^2 - 2\widetilde{\omega^2}\right)\right]$$

where

$$\begin{aligned} y_{j,\cdot}^2 &= \frac{1}{S}\sum_{s=0}^{S-1}\left(Y_{\widetilde{\delta}(j+\frac{s}{S})} - Y_{\widetilde{\delta}(j-1+\frac{s}{S})}\right)^2, \; j = 1, ..., \widetilde{n} \\ \widetilde{\omega^2} &= \exp\left\{\log\left(\widehat{\omega^2}\right) - \widehat{\theta}_n^*/RV\right\} \\ \widetilde{n} &= \left\lfloor 1/\widetilde{\delta}\right\rfloor \end{aligned}$$

and where $\widehat{\theta}_n^*$ is a consistent estimator of integrated variance IV_X. We take $\widehat{\theta}_n^*$ to be the Two-Scale Realized Volatility estimator $\widehat{\theta}_n$. This estimator requires to choose $\widetilde{\delta}$ and S. We use the same choice as Barndorff-Nielsen et al. (2008a) do, for real and simulated data. This choice is $S = n^{1/2}$ and $\widetilde{\delta} = n^{-1/2}$. Estimator $\widetilde{\omega^2}$ corrects small sample bias in $\widehat{\omega^2}$. With large number of observations, there is no difference between the two estimators in practice, but we keep the version of Barndorff-Nielsen et al. (2008a) anyway. Thus, the following alternative estimator of V is constructed

$$\widehat{V}_a = c\frac{4}{3}\widehat{IQ}_{BNHLS} + 8c^{-2}\left[\widetilde{\omega^2}\right]^2.$$

This estimator is consistent for V in the presence of i.i.d. noise.

When noise is autocorrelated, the estimator $\widehat{IQ}_{BNHLS}$ is inconsistent, but the sign of the bias depends on the exact parameters of the model. We now give heuristic explanation about what behavior can be expected of $\widehat{IQ}_{BNHLS}$ in the presence of autocorrelated noise. To simplify the exposition, notice that $y_{j,\cdot}^2$ can be thought of as simply one low frequency return squared. This is because $y_{j,\cdot}^2$ is an average over returns that are very highly correlated given large overlaps in time they have. Also, use $\widehat{\omega^2}$ instead of $\widetilde{\omega^2}$. Consider the building block of $\widehat{IQ}_{BNHLS}$, for $j = 1$,

$$y_{1,\cdot}^2 - 2\widehat{\omega^2} \approx \left(Y_{\frac{1}{n}+\frac{1}{\sqrt{n}}} - Y_{\frac{1}{n}}\right)^2 - \frac{1}{n}\sum_{i=1}^{n}\left(Y_{\frac{i}{n}} - Y_{\frac{i-1}{n}}\right)^2.$$

Suppose $\epsilon \perp X$ and that low frequency noise is uncorrelated. Then,

$$\mathrm{E}\left(y_{1,\cdot}^2 - 2\widehat{\omega^2}\right) \approx \mathrm{E}\left(X_{\frac{1}{n}+\frac{1}{\sqrt{n}}} - X_{\frac{1}{n}}\right)^2 - \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left(Y_{\frac{i}{n}} - Y_{\frac{i-1}{n}}\right)^2 + 2\mathrm{E}\left(\epsilon_{\frac{i}{n}}\epsilon_{\frac{i-1}{n}}\right).$$

The middle term is of smallest order and so can be ignored. Notice that last term is negative in practice, so $\mathrm{E}\left(y_{1,\cdot}^2 - 2\widehat{\omega^2}\right)$ becomes smaller. If noise is not too large, $y_{1,\cdot}^2 - 2\widehat{\omega^2}$ is biased towards zero. Thus, the final estimator of IQ has a negative bias.

Results are represented in terms of coverage probabilities of 95% two-sided, left-sided, and right-sided confidence intervals for IV_X. Table B.1 contains the larger noise-to-signal case $\xi^2 = 0.001$, and Table B.2 contains results for the smaller noise-to-signal case $\xi^2 = 0.0001$. We see that the subsampling estimator performs well in all scenarios. $\widehat{V}_a$ performs well in the scenario it is designed for, which is the uncorrelated noise case. As the correlation increases, estimated values of $\widehat{V}_a$ decrease, resulting in undercoverage. For the $\xi^2 = 0.001$ scenario, two sided coverage probabilities of $\widehat{V}_a$ decrease from 0.93 to 0.78 as autocorrelation becomes stronger. For the smaller noise scenario with $\xi^2 = 0.0001$, two sided coverage probabilities decrease from 0.93 to 0.88. This improvement as noise becomes smaller is to be expected, given that $\widehat{V}_a$ is consistent for V when noise is zero.

2.6 Empirical Analysis

This section applies the proposed subsampling method to tick data of AIG, GE, IBM, INTC, MMM, and MSFT stocks obtained from the NYSE TAQ database, and compares it to the estimator $\widehat{V}_a$, which is introduced in the previous section.

We use the whole year of 2006 of transaction prices for AIG, GE, IBM, INTC, MMM, and MSFT stocks obtained from the NYSE TAQ database. Zero returns are removed, as in Aït-Sahalia et al. (2006a). Griffin and Oomen (2008) show that, in Realized Volatility case, this adjustment of data improves precision of estimation. Jumps are also removed,[8] since the additive market microstructure noise model (2.13) does not allow for jumps. There is also an additional issue to consider, which Barndorff-Nielsen et al. (2008b) denote as local trends or "gradual" jumps. These authors notice that the realized kernel, which is the estimator of integrated variance they propose, does not behave well in the presence of these "gradual" jumps. Such episodes occur rarely, but are nonetheless important. Barndorff-Nielsen et al. (2008b) notice that these local trends are associated with high volumes traded, and conjecture that they are due to non-trivial liquidity effects. The authors replace them with one genuine jump, but conclude that they do not have an automatic way of detecting episodes of local trends. The subsampling method proposed in the current paper also is vulnerable to such price behavior. Our strategy to identify these gradual jumps is based on the fact that they should look like genuine jumps on a lower frequency. Therefore, we construct a time series of lower (one minute) frequency data, and remove those lower frequency returns that are larger than seven weekly standard deviations. Such "gradual" jumps occur rarely, and most weeks do not contain any.

In the resulting data set, the average number of trades per week is 20,341 for AIG,

[8] Jumps are identified as deviations of the log-returns that are larger than five standard deviations on a moving window. This is motivated by the thresholding technique of filtering out jumps, first proposed by Cecilia Mancini in a series of papers (e.g., Mancini 2004), see also Aït-Sahalia and Jacod (2007).

35,361 for GE, 23,657 for IBM, 51,092 for INTC, 15,642 for MMM, and 45,646 for MSFT. The returns of all these stocks display large negative autocorrelation similar to GE in Figure 2.4(b).

The asymptotic variance of the Two Scale Realized Volatility estimator is estimated for each of the 52 weeks in year 2006. As long as the distance between observations is of order $1/n$, the underlying theory can be extended to the non equidistant observations case. Therefore, the estimation is done in tick time, as suggested in Barndorff-Nielsen et al. (2008a) and other authors.

The results are displayed in Figures B.1-B.6 in the appendix B.2, in terms of 95% confidence intervals for integrated variance. Confidence intervals with bars correspond to subsampling method and confidence intervals with lines correspond to the alternative method $\widehat{V}_a$. The TSRV estimate $\widehat{\theta}_n$ is in the center of both confidence intervals by construction. The subsampling confidence intervals for TSRV are usually wider than confidence intervals of the alternative method $\widehat{V}_a$. From our simulations, we conclude this might be due to negative bias of the $\widehat{V}_a$ estimator in the presence of negatively autocorrelated returns. This is because all six stocks have strongly negatively correlated returns, and we know from Section (2.5) that $\widehat{V}_a$ is downward biased in this case. On the other hand, subsampling estimator is immune to autocorrelation. The Figures also show a lot of variability in the estimates of V. This is mainly due to variability of the TSRV estimates, with large estimates of V corresponding to large $\widehat{\theta}_n$ and vice versa. Thus, episodes of high volatility generally correspond to episodes of high volatility of volatility. Though not reported here, these also correspond to weeks with very large numbers of transactions and large volumes traded.

2.7 Conclusion

This paper develops the first automated method for estimating the asymptotic variance of an estimator in high frequency data. The method applies to an important general class of estimators, which include many estimators of integrated variance. The new method can substantially simplify the inference question for an estimator, which has an asymptotic variance that is hard to derive or takes a complicated form. An example of such case is the integrated variance estimator of Aït-Sahalia et al. (2006a), in the presence of autocorrelated heteroscedastic market microstructure noise. There is no alternative inferential method available in the literature in this case.

A question that is yet to be addressed rigorously is a data-driven bandwidth choice. Several choices for the Two Scale Realized Volatility estimator are suggested in the Monte Carlo section.

A very promising extension that will be considered in a future paper is inference for a multivariate parameter. Subsampling naturally produces positive semi-definite estimated variance-covariance matrices, which can be very important for applications. For estimators like Realized Volatility, all the results extend readily to the multivariate case. The real challenge, however, arises due to the additional complications, which are not present in the univariate case. These concern the fact that different stocks do not trade at the same time or so-called asynchronous trading. Also, uncertainty about the observation times becomes much more important in the multivariate context.

Chapter 3

Subsampling and Time Variation in Betas

3.1 Introduction

This paper studies the question of time variability in equity betas. Recent developments in high frequency econometrics allows us to estimate quadratic variation version of the betas in a model-free framework with ultra high frequency and asynchronously observed data from NYSE. Due to the market microstructure noise in this data, estimators of beta can have complicated expressions of the asymptotic variance, in which case it is convenient to use an automatic inference method to implement tests on betas. The subsampling method of Chapter 2 cannot be applied directly to the beta estimators, because they do not satisfy the basic requirement of the method, additivity of the variance of beta estimator. On the other hand, we can easily derive the asymptotic variance of a beta estimator by Delta method if we use a multivariate inference method. Therefore, we show how subsampling of Chapter 2 can be implemented for a multivariate estimator. We then proceed to construct a test of betas being constant over time. We implement this test with six stocks on the NYSE over year 2006 with Standard and Poors Depositary Receipts (SPIDERS) as a proxy for the market factor. We conclude that the traditional 5, 15, or 20 minute based estimators can easily detect significant time variation over the whole year 2006 and also over some quarters for some stocks, but ultra high frequency estimators can detect significant time variation in betas for every quarter and every stock we consider.

The remainder of this paper is organized as follows. Section 3.2 introduces beta and shows how estimation and inference can be done with moderate and high frequency data. Section 3.3 shows how subsampling can be implemented to perform inference on a multivariate parameter. Section 3.4 implements this methodology on high frequency data to test the hypothesis that beta is constant over time. Section 3.6 concludes.

3.2 Realized Beta

Denote X to be the log-price process of the market portfolio, and Y to be a log-price process of an individual stock. Suppose they both follow a Brownian semimartingale process,

$$\begin{aligned} dX_t &= \mu_t^x dt + \sigma_t^x dW_t^x, \\ dY_t &= \mu_t^y dt + \sigma_t^y dW_t^y, \end{aligned} \tag{3.1}$$

over k intervals of length one, i.e., $t \in [i-1, i)$, $i = 1, ..., k$. In above, μ^x and μ^y are drift processes, σ^x and σ^y are volatility processes, and W^x and W^y are standard Brownian Motion processes with $\text{Corr}(W_t^x, W_t^y) = \rho_t$. In continuous time, a natural measure of the variability of the process X over some interval $[i-1, i)$ is its quadratic variation,

$$\langle X, X \rangle_i = \int_{i-1}^{i} (\sigma_t^x)^2 \, dt,$$

and similarly covariation

$$\langle X, Y \rangle_i = \int_{i-1}^{i} \sigma_t^x \sigma_t^y \rho_t dt$$

measures the covariability of X and Y. Using these, we can define a quadratic variation version of the beta over interval $[i-1, i)$ as

$$\beta_i := \frac{\langle X, Y \rangle_i}{\langle X, X \rangle_i}. \tag{3.2}$$

This quantity has been estimated in a low frequency framework by Andersen et al. (2004) and Andersen et al. (2005). Barndorff-Nielsen and Shephard (2004) estimate the same quantity with ultra high frequency data after accounting for market microstructure noise. We consider both approaches.

We refer to this quadratic variation based beta as just beta from now on. Our goal is to construct a test for time variation in betas. To do that, we need to use discrete observations to estimate beta and conduct inference. We consider two different observation schemes: 1) no noise and synchronous observations, and 2) noisy and asynchronous observations.

3.2.1 Estimation of beta: noise-free and synchronous data

The first observation scheme is appropriate for moderate frequencies, such as 5 or 15 minutes. In this case, it is reasonable to assume that we observe efficient stock prices X without error. In general, obtaining 5 or 15 minute calendar time data entails some interpolation, but the effect of this is negligible. Therefore, we can reasonably safely assume that data is both free of microstructure noise and synchronously observed. Denote by n the number of observations in each time period, so that the distance between observations is $1/n$. In this relatively simple scenario we can estimate quadratic variation by realized variance

$$[X,X]_i = \sum_{j=1}^{n} \left(X_{(i-1)n+j} - X_{(i-1)n+j-1}\right)^2$$

and quadratic covariation by realized covariance,

$$[X,Y]_i = \sum_{j=1}^{n} \left(Y_{(i-1)n+j} - Y_{(i-1)n+j-1}\right)\left(Y_{(i-1)n+j} - Y_{(i-1)n+j-1}\right).$$

These can be used to calculate the realized beta as explored by Andersen et al. (2004), Andersen et al. (2005), and Barndorff-Nielsen and Shephard (2004),

$$\widehat{\beta}_i^{RV} := \frac{[X,Y]_i}{[X,X]_i}.$$

It is consistent for the true beta as $1/n \to 0$, under the assumption of no noise and synchronous observations. Asymptotic distribution of realized covariation matrix was first derived in Barndorff-Nielsen and Shephard (2004). We have

$$\sqrt{n}\begin{pmatrix} [X,X]_i - \langle X,X\rangle_i \\ [X,Y]_i - \langle X,Y\rangle_i \end{pmatrix} \Rightarrow N\left(0,\Psi_i\right)$$

where

$$\Psi_i = \int_{i-1}^{i} \begin{pmatrix} 2\sigma_x^4(u) & 2\sigma_x^3(u)\,\sigma_y(u)\,\rho(u) \\ 2\sigma_x^3(u)\,\sigma_y(u)\,\rho(u) & \sigma_x^2(u)\,\sigma_y^2(u)\left(1+\rho^2(u)\right) \end{pmatrix} du. \tag{3.3}$$

By Delta method, provided $\langle X,X\rangle_i > 0$ for $i = 1,..,k$,

$$\sqrt{n}\left(\widehat{\beta}_i^{RV} - \beta_i\right) \Rightarrow N\left(0, \langle X,X\rangle_i^{-2}\begin{pmatrix} -\beta_i & 1 \end{pmatrix}\Psi_i\begin{pmatrix} -\beta_i \\ 1 \end{pmatrix}\right).$$

From Barndorff-Nielsen and Shephard (2004), we know that the asymptotic variance of $\widehat{\beta}_i^{RV}$ can be estimated by[1]

$$[X,X]_i^{-2}\left(\sum_{j=1}^{n} k_{i,j}^2 - \sum_{j=1}^{n-1} k_{i,j}k_{i,j+1}\right) \tag{3.4}$$

where

$$\begin{aligned} k_{i,j} &= \left(X_{(i-1)n+j} - X_{(i-1)n+j-1}\right)\left(Y_{(i-1)n+j} - Y_{(i-1)n+j-1}\right) \\ &\quad -\widehat{\beta}_i^{RV}\left(X_{(i-1)n+j} - X_{(i-1)n+j-1}\right)^2. \end{aligned}$$

Joint distribution of estimated betas for different times can be obtained from marginals, since asymptotic distributions of $\sqrt{n}\left(\widehat{\beta}_i^{RV} - \beta_i\right)$ are independent for any $i \neq j$ (see e.g. Mykland and Zhang, 2006).

Gonçalves and Meddahi (2007) have proposed a multivariate bootstrap that performs better than estimator in (3.4) in finite samples. Thus, it can be used as an alternative inference method for this sampling scheme. As it stands, their procedure is not applicable to the second sampling scheme we will now consider, although a modification might exist which can be used.

3.2.2 Estimation of beta: noisy and asynchronously observed data

We now turn to the second observation scheme. If we want to make use of the full record of transaction prices, it becomes important to account both for asynchronous observations across stocks and market microstructure noise. The latter is typically modeled as an additive measurement error. It means that instead of having observations on X and Y, we have observations on $X+\epsilon^x$ and $Y+\epsilon^y$. We assume noise (ϵ^x, ϵ^y)

[1]The meaning of g here and in BN-S (2004) coincide, but Ψ_i is defined slightly differently.

is independent of the efficient price (X, Y), it is stationary, it has more than four moments, and it is exponentially α-mixing when viewed as a process in observation times in $\mathcal{G}^{(i)}$, which is defined below.

Number of observations will in general be different for each interval and each stock, so additional notation is needed. For the i^{th} interval, we have $n^{x,(i)}$ return observations on $X + \epsilon^x$ and $n^{y,(i)}$ return observations on $Y + \epsilon^y$ at non-equidistant, deterministic times. Let $\mathcal{G}^{x,(i)}$ be a set that contains transaction/observation times for $X + \epsilon^x$ on day i,

$$\mathcal{G}^{x,(i)} := \left\{t_0^{x,(i)}, t_1^{x,(i)}, t_2^{x,(i)}, ..., t_{n^{x,(i)}}^{x,(i)}\right\}, \ i = 1, .., k$$

where $t_j^{x,(i)}$ is the j^{th} transaction of X on the i^{th} day. Define $\mathcal{G}^{y,(i)}$ similarly. We suppress the dependence of $\mathcal{G}_i^x$ and every its element $t_j^{x,(i)}$ on $n^{x,(i)}$. This triangular structure arises due to infill asymptotics, i.e., asymptotics as the time distance between any two observations shrinks. In particular, we will require, for each interval i (see Zhang 2008),

$$\begin{aligned}
\sup_j \left|t_j^{x,(i)} - t_{j-1}^{x,(i)}\right| &= O\left(\left(n^{x,(i)} + n^{y,(i)}\right)^{-1}\right) \text{ and} \\
\sup_j \left|t_j^{y,(i)} - t_{j-1}^{y,(i)}\right| &= O\left(\left(n^{x,(i)} + n^{y,(i)}\right)^{-1}\right).
\end{aligned}$$

Observations need to be synchronized between X and Y before proceeding. To do this, we use the Refresh Time idea of Barndorff-Nielsen et al. (2008c). This means creating a new set of "observation times" $\mathcal{G}^{(i)}$ with elements $t_j^{(i)}$ as follows. The first element $t_0^{(i)}$ is the first time both stocks have traded, $t_0^{(i)} = \max\left\{t_0^{x,(i)}, t_0^{y,(i)}\right\}$. After that, next element $t_j^{(i)}$ is the earliest moment when both stocks have again traded at least once, i.e., $t_j^{(i)}$ is the maximum of $\min\left\{t \in \mathcal{G}^{x,(i)} : t > t_{j-1}^{(i)}\right\}$ and $\min\left\{t \in \mathcal{G}_i^y : t > t_{j-1}^{(i)}\right\}$. Next, we do not actually have transactions/observations for for X and Y at each time in $\mathcal{G}^{(i)}$, so we obtain a new set of data, X^o and Y^o, by previous-tick interpolation to times in $\mathcal{G}^{(i)}$.[2] Denote by $n^{(i)}$ the number of constructed returns on the i^{th} day (i.e., $n^{(i)}$ is defined by $\left|\mathcal{G}^{(i)}\right| = n^{(i)} + 1$).

Beta can be estimated using very high frequency data using Two Scale estimator of Zhang et al. (2005) or Aït-Sahalia (2006a), Multi Scale estimator of Zhang (2006), Realized Kernels of Barndorff-Nielsen et al. (2008a), or pre-averaging estimator of Jacod et al. (2007).[3] We will use the Two Scale estimator of Aït-Sahalia et al. (2006a), which was extended to multivariate setting by Zhang (2008), to obtain the following estimator of beta,

$$\widehat{\beta}_i^{AMZ} = \frac{\widehat{\langle X,Y\rangle}_i^{AMZ}}{\widehat{\langle X,X\rangle}_i^{AMZ}}, \; i = 1, ..., k. \tag{3.5}$$

In what follows, it is more convenient to use notation n_i instead of $n^{(i)}$. From the joint asymptotic distribution

$$n_i^{1/6}\left(\left(\begin{array}{c}\widehat{\langle X,X\rangle}_i^{AMZ}\\ \widehat{\langle X,Y\rangle}_i^{AMZ}\end{array}\right) - \left(\begin{array}{c}\langle X,X\rangle_i\\ \langle X,Y\rangle_i\end{array}\right)\right) \Rightarrow N\left(0, \Sigma_i^{AMZ}\right), \; i = 1, ..., k,$$

we obtain the asymptotic distribution for realized beta by the Delta method (provided $\langle X,X\rangle_i > 0$),

$$n_i^{1/6}\left(\widehat{\beta}_i^{AMZ} - \beta_i\right) \Rightarrow N\left(0, V_i^{AMZ}\right), \; i = 1, ..., k$$

where

$$V_i^{AMZ} = \langle X,X\rangle_i^{-2}\left(\begin{array}{cc}-\beta_i & 1\end{array}\right)\Sigma_i^{AMZ}\left(\begin{array}{c}-\beta_i\\ 1\end{array}\right).$$

The exact expression of Σ_i^{AMZ} is rather complicated, and the reader can find it in

[2]For practical implementation, the algorithm in Palandri (2006) representing this synchronization process is useful.

[3]See Kinnebrock and Podolskij (2008) for a multivariate generalisation of Jacod et al. (2007) and a related paper Podolskij and Vetter (2007). See Barndorff-Nielsen et al. (2008c) for a multivariate generalisation of Barndorff-Nielsen et al. (2008a).

Zhang (2008). We do not need the exact expression for estimation because we will use subsampling to estimate Σ_i^{AMZ}, see Section 3.3. Asymptotic distributions of $n_i^{1/6}\left(\widehat{\beta}_i^{AMZ} - \beta_i\right)$ are again independent across periods i. Next, we show how to construct a test that the true beta is constant across time.

3.2.3 Testing for constant betas

Using the joint asymptotic distribution of betas across k time periods, we can construct a Chi-square test for the true betas being constant across these time periods. The construction will be based on $\widehat{\beta}^{AMZ}$, but exactly the same idea can be used to construct the test based on $\widehat{\beta}^{RV}$. Define

$$\begin{aligned}\widehat{\beta}^{AMZ} &= \left(\begin{array}{cccc}\widehat{\beta}_1^{AMZ} & \widehat{\beta}_2^{AMZ} & \ldots & \widehat{\beta}_k^{AMZ}\end{array}\right)' \text{ and} \\ \beta &= \left(\begin{array}{cccc}\beta_1 & \beta_2 & \ldots & \beta_k\end{array}\right)'.\end{aligned}$$

Given that asymptotic distributions of $\widehat{\beta}_i^{AMZ} - \beta_i$ are independent across time periods $i = 1, 2, ..., k$, we obtain the joint asymptotic distribution from the marginal ones. The asymptotic variance covariance matrix will be diagonal. We have

$$\Phi\left(\widehat{\beta}^{AMZ} - \beta\right) \Rightarrow N\left(0, V^{AMZ}\right)$$

where

$$\begin{aligned}V^{AMZ} &= diag\left(V_1^{AMZ}, V_2^{AMZ}, ..., V_k^{AMZ}\right) \\ \Phi &= diag\left(n_1^{1/6}, n_2^{1/6}, ..., n_k^{1/6}\right).\end{aligned}$$

We are interested in testing the hypothesis that true beta is constant over time

$$H_0 : \beta_1 = ... = \beta_k, \text{ versus } H_1 : \beta_i \neq \beta_j \text{ for some } i \text{ and } j.$$

One way to construct a test is to use the sum of squared differences $\widehat{\beta}_i^{AMZ} - \widehat{\beta}_1^{AMZ}$ for $i = 2, ..., k$, properly standardized. For this purpose, introduce the following $k-1$ by k matrix

$$\Delta = \begin{pmatrix} -1 & 1 & 0 & 0 & ... & 0 \\ -1 & 0 & 1 & 0 & ... & 0 \\ -1 & 0 & 0 & 1 & ... & 0 \\ ... & ... & ... & ... & ... & ... \\ -1 & 0 & 0 & 0 & ... & 1 \end{pmatrix}.$$

This matrix can be used to construct a vector of length $k-1$ containing all differences,

$$\Delta\Phi\left(\widehat{\beta}^{AMZ} - \beta\right) = \Phi \begin{pmatrix} \widehat{\beta}_2^{AMZ} - \widehat{\beta}_1^{AMZ} - (\beta_2 - \beta_1) \\ \widehat{\beta}_3^{AMZ} - \widehat{\beta}_1^{AMZ} - (\beta_3 - \beta_1) \\ ... \\ \widehat{\beta}_k^{AMZ} - \widehat{\beta}_1^{AMZ} - (\beta_k - \beta_1) \end{pmatrix} \Rightarrow N\left(0, \Delta V^{AMZ}\Delta'\right).$$

By continuous mapping theorem, we have

$$\left(\widehat{\beta}^{AMZ} - \beta\right)' \Phi\Delta' \left(\Delta V^{AMZ}\Delta'\right)^{-1} \Delta\Phi\left(\widehat{\beta}^{AMZ} - \beta\right) \Rightarrow \mathcal{X}_{k-1}^2.$$

Notice that under the null, $\Delta\Phi\left(\widehat{\beta}^{AMZ} - \beta\right) = \Delta\Phi\widehat{\beta}^{AMZ}$. We can now take any consistent estimator $\widehat{V}^{AMZ}$ and construct the following test statistic,

$$T = \left(\widehat{\beta}^{AMZ}\right)' \Phi\Delta' \left(\Delta\widehat{V}^{AMZ}\Delta'\right)^{-1} \Delta\Phi\widehat{\beta}^{AMZ},$$

which has

$$T \Rightarrow \mathcal{X}^2_{k-1} \text{ under } H_0.$$

Under the alternative H_1, this test statistic T diverges to infinity, meaning that the test is consistent. Also, estimate of the beta of the first interval was used for centering, but it can be shown that the resulting test statistic is invariant to the choice of centering.

3.3 Multivariate Subsampling

The aim of this section is to generalize the ideas of Chapter 2 to the multivariate framework, thus producing an automatic, positive semi-definite estimator of the asymptotic variance-covariance matrix. In the next section, this method will be used to estimate V^{AMZ} and implement the test of time invariant betas. To simplify notation, consider the case of one interval only, $k = 1$, and drop the interval subscript i. We seek to estimate matrix Σ in

$$\tau_n \left(\widehat{\theta} - \theta\right) \Rightarrow N\left(0, \Sigma\right) \tag{3.6}$$

where τ_n is a known rate of convergence when n observations are used. The rate of convergence is assumed to be the same for all elements of $\widehat{\theta}$. In the application later, we will choose some interval i, take n to be the number of synchronized observations n_i, choose

$$\theta = \begin{pmatrix} \langle X, X \rangle \\ \langle X, Y \rangle \end{pmatrix},$$

estimate θ using synchronized data X^o and Y^o, and then use Delta method to do inference for β. Note that none of the estimators of β considered so far can be subsampled directly. That is because one of the key requirements of the subsampling method in Chapter 2 is additivity over time of the asymptotic variance, and estimators of β considered in the previous section do not have this property.

The subsampling method of Chapter 2 is based on a series of longer blocks of observations, m returns in each block, as well as a series of shorter blocks of observations, J returns in each block, $J < m < n$, see Figure 3.1. In Figure 3.1, observations are equidistant, which we do not assume of course. Denote $\mathcal{A}$ to be a set containing some observation times, and $\widehat{\theta}(\mathcal{A})$ to be an estimator $\widehat{\theta}$ calculated using observations at times in $\mathcal{A}$. Using this notation, the subsampling estimator of the asymptotic variance-covariance matrix Σ is

$$\widehat{\Sigma}_{sub} = \frac{J}{n}\frac{1}{K}\sum_{l=1}^{K}\tau_n^2\left(\frac{n}{J}\widehat{\theta}_l^{short} - \frac{n}{m}\widehat{\theta}_l^{long}\right)\left(\frac{n}{J}\widehat{\theta}_l^{short} - \frac{n}{m}\widehat{\theta}_l^{long}\right)' \tag{3.7}$$

where

$$\begin{aligned}
\widehat{\theta}_l^{long} &= \widehat{\theta}\left(\left\{t_{(l-1)s+1}, t_{(l-1)s+2}, \ldots, t_{(l-1)s+m+1}\right\}\right)\\
\widehat{\theta}_l^{short} &= \widehat{\theta}\left(\left\{t_{(l-1)s+1}, t_{(l-1)s+2}, \ldots, t_{(l-1)s+J+1}\right\}\right)\\
K &= \left\lceil \frac{n-m}{s} + 1 \right\rceil,
\end{aligned}$$

K is the number of subsamples, and s stands for "shift", i.e., by how many observations to roll the window to obtain the next subsample. Thus, it controls the amount of overlap between the subsamples. The smallest s is 1 and it corresponds to the maximum overlap and largest number of subsamples; Figure 3.1 is drawn for this case. This is also the most efficient choice. However, it can be very computationally intensive in practice, so a larger s can also be used at the expense of less efficient, but nevertheless consistent $\widehat{\Sigma}_{sub}$. From the definitions of $\widehat{\theta}_l^{short}$ and $\widehat{\theta}_l^{long}$ above, we can see that longer and shorter subsamples start at the same time. This case is less involved to write down, but in practice the case drawn in Figure 3.1 is better, i.e., both subsamples are centered at the same time. For this case, shorter subsample should start at $t_{(l-1)s+\lfloor (m-J)/2 \rfloor+1}$ and not $t_{(l-1)s+1}$. Also, notice that the formula in (3.7) simplifies to one in Chapter 2 for univariate estimator and no overlap case (i.e., $s = m$ and $K = \lceil n/m \rceil$).

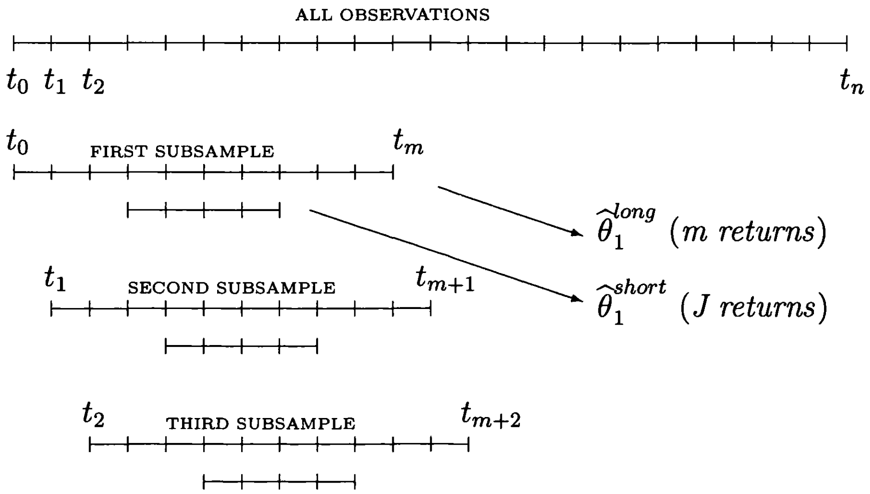


Figure 3.1: *The Subsampling Scheme of Chapter 2*

It is easy to see that $\widehat{\Sigma}_{sub}$ is positive semi-definite by construction. This avoids any risk of length of estimated confidence intervals for $\widehat{\beta}$ (or other continuous functionals of elements of $\widehat{\theta}$) being negative.

3.4 Empirical Analysis

In this section we implement the test of constant betas based on $\widehat{\beta}^{RV}$ and $\widehat{\beta}^{AMZ}$. To implement the test, we will need estimated variances of $\widehat{\beta}^{RV}$ and $\widehat{\beta}^{AMZ}$. To this end, we use BN-S (2004) estimator of the variance-covariance matrix of $\widehat{\beta}^{RV}$, and the estimator of variance-covariance matrix of $\widehat{\beta}^{AMZ}$ obtained by subsampling as in Section 3.3. We calculate $\widehat{\beta}^{AMZ}$ using all data, but for $\widehat{\beta}^{RV}$ we need to choose some lower sampling frequency. We choose three frequencies that are popular in practice, 5, 15, and 20 minutes, and denote the resulting estimators as $\widehat{\beta}^{RV}_{5min}$, $\widehat{\beta}^{RV}_{15min}$, and $\widehat{\beta}^{RV}_{20min}$.

3.4.1 The Data

We use high frequency transactions data on six individual stocks. They are American International Group, Inc. (listed under the ticker symbol AIG), General Electric Co. (GE), International Business Machines Co. (IBM), Intel Co. (INTC), Minnesota Mining and Manufacturing Co. (MMM), and Microsoft Co. (MSFT). To proxy for a market portfolio, we use Standard and Poors Depositary Receipts (SPIDERS for short, ticker symbol SPY), which are an Exchange Traded Fund set up to mimic the movements of the Standard and Poor's 500 Composite Stock Price Index. Our data covers the whole year 2006 and is obtained from the NYSE Trade and Quote database.

We clean the data as follows. We apply time filter 9:30 - 16:00. We retain all satellite markets. Where multiple transactions per second are recorded, we take the first one. Where repeated times are recorded, we take the average. Next, we delete bounce backs, jumps, as well as gradual jumps as follows. Bounce backs are most likely to result from data mistakes, such as incorrect time record, so as a first step we identify bounce backs among prices and delete them. We define bounce backs as two consecutive price changes of the opposite sign, where each of the two price changes is larger, in absolute value, than five standard deviations of the observed price over a moving window of 500 transactions. Next, we remove jumps using the thresholding methodology of Mancini (2004). In other words, we set those returns that are larger than some threshold to zero. The threshold for this purpose is defined as five standard deviations of the observed price, and is calculated over a moving window of one day. Finally, we remove gradual jumps. Barndorff-Nielsen et al. (2008b) discuss the fact that Realized Kernels do not behave well when price only rises (falls) over some period of time. Two Scale estimator is similarly not robust to gradual jumps, so we also have to deal with them. Barndorff-Nielsen et al. (2008b) define gradual jumps as relatively long periods containing only price increases or only price decreases. They then replace the returns of this period with one single jump. We define gradual jumps as at least 5 minutes long interval containing only price increases (or decreases), provided the total price change exceeds a threshold of five standard deviations of the observed returns. Gradual jumps are replaced with a zero return. The threshold is recalculated over a moving window of 5 days. All window lengths mentioned in the cleaning procedure are average ones; windows are fixed in terms of number of transactions so as to achieve an average target of calendar time period over the year.

In order to calculate realized betas, we need to synchronize the data (see Section 3.2.2). To maximize the information used, we synchronize data in pairs only. For example, to estimate beta of INTC, we synchronize INTC with SPY; to estimate beta of MMM, we synchronize MMM with SPY. Therefore, different transformations of the original SPY data is used to calculate different betas.

3.4.2 Results

We start by analyzing the data. Table C.1 contains some summary statistics of the data before synchronization: transactions per week, estimates of the noise variance, noise-to-signal ratio, and autocorrelations of returns at first three lags. First autocorrelations are all large and negative, which is typical of noisy data and unlikely to arise from Brownian Semimartingale. Second autocorrelations are all positive, some are large. Alternating signs of autocorrelations indicate that the main component of the noise is bid-ask bounce. In fact, if we removed all zero returns, the remaining data would display very persistent autocorrelation with alternating signs (see figure with autocorrelations in Chapter 2, this has been also noted in e.g. Griffin and Oomen 2005). In full data set with zero returns, this effect is attenuated because switching times of bid and ask are random. Third autocorrelations are of different signs and small. The estimates of the noise variance (columns 2 and 3 in Table C.1) are very small, and in fact several orders of magnitude smaller than Hansen and Lunde (2005) estimates for year 2004. For example, the simplest estimator of the noise variance is

$$\widehat{\omega}^2 = [X, X] / 2n.$$

Our estimate for INTC in 2006 is $0.518 \cdot 10^{-7}$, while Hansen and Lunde (2005) report this number for 2004 to be $0.46 \cdot 10^{-3}$. Apart from the obvious fact that years are different, there are also important differences in methodology. We calculate $\widehat{\omega}^2$ using the whole year, they calculate it every day and report the annual average. Data cleaning can also be an important source of differences. The more data is cleaned, the smaller estimate of the noise variance we would expect. However, it does not seem that these reasons can explain differences of such magnitude.

Table C.2 contains the same summary statistics for the cleaned data. As long as there is any asynchronicity in the observations, number of synchronized observations will be smaller. We can see the reduction of the transactions per week by comparing first columns of Table C.1 and C.2. Noise variances are larger as measured by $\widehat{\omega}^2$, but we can easily verify this is purely due to larger finite-sample bias caused by smaller number of observations. In particular, the bias-adjusted estimators

$$\widehat{\omega}^2 = \left([X, X] - \widehat{\langle X, X \rangle}^{AMZ} \right) / 2n$$

are the same with and without synchronization. Autocorrelations are smaller, which is due to frequency being lower.

Figure C.1 contains volatility signature plots for each individual stock (plots of realized variance agains the frequency used in its calculation), as well covariance signature plots (plots of realized covariance against the frequency). Volatility signature plots show a large increase for highest frequencies, consistent with the additive noise model where bias explodes as we sample more and more frequently. On the other hand, realized covariances display the so-called Epps effect due to Epps (1979), i.e., they tend to zero as the frequency increases. Zhang (2008) analytically characterizes this bias for realized covariance based on previous-tick interpolated prices (Refresh Times is a special case of her approach). Therefore, neither Realized Variance, nor Realized Covariance should be calculated using the highest frequencies. On the other hand, the Two Scale estimator, while using all the synchronized data, cancels both the effect of noise and asynchronous observations and is consistent.

Figures C.2 - C.4 show plots of estimated betas using $\widehat{\beta}_{5min}^{RV}$ and $\widehat{\beta}^{AMZ}$ together with 95% confidence intervals. We see that beta is estimated much more precisely using all the data. The two parameters in $\widehat{\beta}^{AMZ}$ were chosen as follows. G_1 was set to the number of ticks as to correspond to 5 minutes on average. G_2 was set to 3, given that there is no evidence of autocorrelations at larger lags. The two parameters of the subsampling scheme were set to $m = 20G_1$ and $J = 5G_1$. Estimates and confidence intervals for $\widehat{\beta}_{15min}^{RV}$ and $\widehat{\beta}_{20min}^{RV}$ are not shown, but they have much longer confidence intervals than $\widehat{\beta}_{5min}^{RV}$.

Table C.3 contains the results of the test for constant betas. In particular, it contains values of the test statistics with corresponding p-values in parenthesis. The null hypothesis is that the true beta is constant over some time period. We implement the test for five different time periods: the whole year 2006, and each quarter separately. Four different tests are implemented based on four estimators: $\widehat{\beta}_{5min}^{RV}, \widehat{\beta}_{15min}^{RV}, \widehat{\beta}_{20min}^{RV}$ and $\widehat{\beta}^{AMZ}$. Roughly speaking, the results of the tests can be anticipated by looking at the figures with point estimates and their confidence intervals. We see that the null hypothesis of beta being constant over the whole year can be rejected using a test based on any of the four estimators. For shorter periods, answer varies depending on the stock and the exact time period. The test based on $\widehat{\beta}^{AMZ}$ can reject the null, at 5% level of significance, for any of scenarios considered, except it has a p-value of 0.057 for GE Q1. The test based on $\widehat{\beta}_{5min}^{RV}$ rejects the null for fewer cases. The test based on $\widehat{\beta}_{15min}^{RV}$ fails to reject the null for roughly half of quarters-based cases, and the test based on $\widehat{\beta}^{RV}_{20min}$ fails to reject the null for most of quarters-based cases.

We conclude that there is enough evidence against betas being constant across the whole year, but high frequency data can uncover significant variation in betas over shorter intervals.

3.5 Simulation Study

This section examines the finite sample properties of the subsampling method with irregular and asynchronous observations.

Efficient log-price of each of the two stocks follows a univariate Heston (1993) model:

$$\begin{aligned} dX_t^{(i)} &= \left(\alpha_1 - v_t^{(i)}/2\right) dt + \sigma_t^{(i)} dW_t^{(i)} \\ dv_t^{(i)} &= \alpha_2 \left(\alpha_3 - v_t^{(i)}\right) dt + \alpha_4 \left(v_t^{(i)}\right)^{1/2} dB_t^{(i)}, \ i = 1, 2 \end{aligned}$$

where $v_t^{(i)} = \left(\sigma_t^{(i)}\right)^2$, $W_t^{(i)}$ and $B_t^{(i)}$ are independent Brownian Motions. The parameters of the univariate efficient log-price process are chosen to be the same as in Zhang et al. (2005). They are $\alpha_1 = 0.05$, $\alpha_2 = 5$, $\alpha_3 = 0.04$, and $\alpha_4 = 0.5$ (the same for $i = 1, 2$). Correlation of the two processes is obtained by setting $Corr\left(W_t^{(1)}, W_t^{(2)}\right) = \varrho$, with ϱ taking values 0, 0.25 , 0.50, and 0.75 across different experiments. In this model, the beta for the i^{th} period is

$$\beta_i = \varrho \int_{i-1}^{i} \sigma_u^{(1)} \sigma_u^{(2)} du \Bigg/ \int_{i-1}^{i} \left(\sigma_u^{(1)}\right)^2 du. \tag{3.8}$$

Microstructure noise is simulated as a normally distributed white noise with variance $\xi^2 IQ^{(1)}$, where ξ^2 is a noise-to signal ratio taking values 0, 0.001, and 0.01, and $IQ^{(1)}$ is the integrated quarticity of the first stock (approximated as a Riemann sum of simulated 1 second values of σ_t^2). Since volatility paths are different across simulations, noise variance also varies across simulations and increases with higher volatility of the efficient price. Observed prices are efficient log-prices plus noise.

As a first step, we simulate one week of 1 second synchronous observations (simulation is done via an Euler scheme with one year as a unit of time and one second step length). From these, we take 35,000 irregular and asynchronous observations for each stock as follows. We draw a random permutation of all observation times in a week, take the first 35,000 of them, and sort them. Observation times are independent across stocks. Observations are then synchronized using the Refresh Time method, resulting in a random number of observations (usually somewhere around 25,000).

The Two Scale estimator is implemented with exactly the same parameters as in the empirical analysis (Section 3.4). In other words, G_1 is taken so as to correspond to 5 minutes on average (typically around 70), and $G_2 = 3$ (see Section 2.3 for the meaning of these parameters). Subsampling parameters are taken to be $J = 5G_1$ and $m = 20G_1$ also as in section 3.4. The resulting coverage probabilities for betas are as follows.

	$\varrho = 0$	$\varrho = 0.25$	$\varrho = 0.5$	$\varrho = 0.75$
$\xi^2 = 0$	0.943	0.926	0.932	0.914
$\xi^2 = 0.001$	0.939	0.927	0.925	0.924
$\xi^2 = 0.01$	0.938	0.925	0.921	0.922

Table 3.1: 95% *coverage probabilities of beta, where beta is estimated using Two Scale estimators as in (3.5). Confidence intervals of the Two Scale estimator are calculated using the asymptotic variance estimated by subsampling.* ξ^2 *is the noise to signal ratio.* ϱ *controls the realizations of beta, see (3.8). Number of simulations is 2000.*

There is some undercoverage, but otherwise subsampling method seems to work reasonably well in finite samples. An interesting finding (not documented here) was that the Two Scale estimator, even in the univariate case, was downward biased with irregular observations when $G_1 = 1$, while it was unbiased for larger values of G_1. This means that the same method that can deal with autocorrelated noise ($G_1 > 1$)

can decrease higher order biases due to irregular observations. Good finite sample properties of the underlying estimator are crucial for the subsampling method to deliver good estimates of the asymptotic variance.

3.6 Conclusion

This paper studies the question of time variability in equity betas. Recent developments in high frequency econometrics allows us to estimate quadratic variation version of the betas in a model-free framework with ultra high frequency and asynchronously observed data from NYSE. Due to the market microstructure noise in this data, estimators of beta can have complicated expressions of the asymptotic variance, in which case it is convenient to use an automatic inference method to implement tests on betas. We show now the multivariate version of the subsampling method of Chapter 2 can be used to achieve this. We then proceed to construct a test of betas being constant over time. We implement this test with six stocks on the NYSE over year 2006 with Standard and Poors Depositary Receipts as a proxy for the market factor. We find that the use of ultra high frequency data allows to detect significant variation of betas over shorter intervals of time than the traditional estimators relying on 5, 15, or 20 minute data.

Bibliography

[1] AÏT-SAHALIA, Y. AND J. JACOD (2008). Estimating the Degree of Activity of Jumps in High Frequency Data. Annals of Statistics, forthcoming.

[2] AÏT-SAHALIA, Y., AND L. MANCINI (2006). Out-of-sample forecasts of Quadratic Variation. Working Paper.

[3] AÏT-SAHALIA, Y., P. MYKLAND, AND L. ZHANG (2005). How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise. Review of Financial Studies, 18, 351-416.

[4] AÏT-SAHALIA, Y., P. MYKLAND, AND L. ZHANG (2006a). Ultra High Frequency Volatility Estimation with Dependent Microstructure Noise. Working Paper.

[5] AÏT-SAHALIA, Y., P. MYKLAND, AND L. ZHANG (2006b). Comments on 'Realized Variance and Market Microstructure Noise,' by P. Hansen and A. Lunde, Journal of Business and Economic Statistics.

[6] ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2001). The distribution of exchange rate volatility. Journal of the American Statistical Association 96, 42 - 55. Correction published in 2003, volume 98, page 501.

[7] ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND J. WU (2004). Realized beta: persistence and predictability, in T. Fomby and D. Terrell (eds.)

Advances in Econometrics: Econometric Analysis of Economic and Financial Time Series in Honor of R.F. Engle and C.W.J. Granger , Volume B, 1-40

[8] ANDERSEN, T.G., T. BOLLERSLEV, F. X. DIEBOLD, AND J. WU (2005). A Framework for Exploring the Macroeconomic Determinants of Systematic Risk. American Economic Review, 95, 398-404.

[9] ANDREOU, E. AND E. GHYSELS (2002). Rolling-Sample Volatility Estimators: Some New Theoretical, Simulations and Empirical Results. Journal of Business and Economic Statistics 20, 363-375

[10] ANDREWS, D. W. K. (1998). Laws of Large Numbers for Dependent Non-Identically Distributed Random Variables. Econometric Theory, Vol. 4, No. 3., 458 - 467

[11] AWARTANI, B., CORRADI, V., AND W. DISTASO (2004). Testing and Modelling Market Microstructure Effects with and Application to the Dow Jones Industrial Average. Working paper.

[12] BACK, K. AND S. BARUCH (2004). Kyle meets Glosten and Milgrom. Econometrica 72(2), 433-465.

[13] BANDI, F. M. AND J. R. RUSSELL (2006a). Separating microstructure noise from volatility. Journal of Financial Economics 79, 655-692.

[14] BANDI, F. M., AND J. R. RUSSELL (2006b). Comment on "Realized Variance and Microstructure Noise" by Peter Hansen and Asger Lunde. Journal of Business and Economic Statistics, 24, 167-173.

[15] BANDI, F. M. AND J. R. RUSSELL (2008). Microstructure noise, realized variance, and optimal sampling. Review of Economic Studies 75, 339–369.

[16] BANDI, F. M., J. R. RUSSELL, AND C. YANG (2007). Realised Volatility Forecasting and Option Pricing. Working paper, Chicago GSB.

[17] BARNDORFF-NIELSEN, O. E., S. E. GRAVERSEN, J. JACOD, AND N. SHEPHARD (2006). Limit theorems for bipower variation in financial econometrics. Econometric Theory, 22, 677-719.

[18] BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A., AND N. SHEPHARD (2006). Subsampling Realized Kernels. Journal of Econometrics, forthcoming.

[19] BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A., AND N. SHEPHARD (2008a). Designing Realised Kernels to Measure the Ex-post Variation of Equity Prices in the Presence of Noise. Econometrica, Vol. 76, No. 6, 1481-1536

[20] BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A., AND N. SHEPHARD (2008b). Realised Kernels in Practice: Trades and Quotes. Econometrics Journal, forthcoming.

[21] BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A., AND N. SHEPHARD (2008c). Multivariate Realised Kernels: Consistent Positive Semi-Definite Estimators of the Covariation of Equity Prices with Noise and Non-synchronous Trading. Working Paper.

[22] BARNDORFF-NIELSEN, O. E. AND SHEPHARD, N. (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society B 64, 253–280.

[23] BARNDORFF-NIELSEN, O. E. AND SHEPHARD, N. (2004). Power and Bipower Variation with Stochastic volatility and Jumps. Journal of Financial Econometrics, Vol. 2, No. 1, 1-37

[24] BARNDORFF-NIELSEN, O. E. AND SHEPHARD, N. (2007). Variation, jumps, market frictions and high frequency data in financial econometrics. Advances in Economics and Econometrics. Theory and Applications, Ninth World Congress, (edited by R. Blundell, P. Torsten, and W. K. Newey), Econometric Society Monographs, Cambridge University Press.

[25] BERK, K. N. (1973). A Central Limit Theorem for m-dependent Random Variables with Unbounded m. The Annals of Probability, Vol. 1, No. 2, 352-354.

[26] BOUND, J., C. BROWN, AND N. MATHIOWETZ (2001). Measurement error in survey data. In the Handbook of Econometrics, Eds. J.J. Heckman and E. Leamer vol. 5. 3705–3843.

[27] CHRISTENSEN, K., R. C. A. OOMEN, AND M. PODOLSKIJ (2008). Realised Quantile-Based Estimation of the Integrated Variance. Available at SSRN: http://ssrn.com/abstract=1085553

[28] CHRISTENSEN, K., M. PODOLSKIJ, AND M. VETTER (2006). Bias-Correcting the Realized Range-Based Variance in the Presence of Market Microstructure Noise. Available at SSRN: http://ssrn.com/abstract=935419

[29] COPELAND, T. E., AND D. GALAI (1983). Information Effects on the Bid-Ask Spread. The Journal of Finance, Vol. 38, No. 5, 1457-1469.

[30] DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. Annals of Statistics 25, 1–37.

[31] DELBAEN, F. AND W. SCHACHERMAYER (1994). A general version of the fundamental theorem of asset pricing. Math. Ann., 300(3): 463-520.

[32] DIEBOLD, F.X. (2006). On Market Microstructure Noise and Realized Volatility. Discussion of Hansen and Lunde (2006).

[33] EPPS, T.W. (1979). Comovements in stock prices in the very short run. Journal of the American Statistical Association, 74, 291-298.

[34] FOSTER, D. P. AND D. B. NELSON (1996). Continuous Record Asymptotics for Rolling Sample Variance Estimators. Econometrica 64, 139-174.

[35] GERETY, M. S., AND H. MULHERIN (1994). Price formation on stock exchanges: The evolution of trading within the day, Review of Financial Studies 6, 23-56.

[36] GHYSELS, E. AND A. SINKO (2006). Volatility Forecasting and Microstructure Noise. Working paper, University of North Carolina.

[37] GLOTER, A., AND J. JACOD (2001). Diffusions with measurement errors. I - Local asymptotic normality. ESAIM: Probability and Statistics 5, 225–242.

[38] GONÇALVES, S. AND N. MEDDAHI (2007). Bootstrapping Realized Multivariate Volatility Measures. Working Paper.

[39] GONÇALVES, S. AND N. MEDDAHI (2008). Bootstrapping Realized Volatility. Forthcoming in Econometrica.

[40] GOTTLIEB, G., AND A. KALAY (1985). Implications of the Discreteness of Observed Stock Prices. The Journal of Finance XL, 135–153.

[41] GRIFFIN, J. AND R. OOMEN (2008). Sampling Returns for Realized Variance Calculations: Tick Time or Transaction Time? Econometric Reviews, Vol. 27, 230-253.

[42] HALL, P., AND C. C. HEYDE (1980). Martingale Limit Theory and Its Applications. Academic Press, New York

[43] HANSEN, P. R., AND A. LUNDE (2006). Realized variance and market microstructure noise (with comments and rejoinder). Journal of Business and Economic Statistics 24, 127–218.

[44] HARRIS, L. (1986). A transaction data study of weekly and intradaily patterns in stock returns, Journal of Financial Economics, Vol. 16, Issue 1, 99-117

[45] HESTON, S. (1993). A Closed-Form Solution for Options With Stochastic Volatility With Applications to Bonds and Currency Options. Review of Financial Studies, 6, 327–343.

[46] JACOD, J. (2006). Asymptotic properties of realized power variations and related functionals of semimartingales. Unpublished paper

[47] JACOD, J. (2007). Statistics and high frequency data. Lecture notes. Séminaire Européen de Statistique 2007.

[48] JACOD, J., Y. LI, P. A. MYKLAND, M. PODOLSKIJ, AND M. VETTER (2007). Microstructure Noise in the Continuous Case: The Pre-Averaging Approach. Stochastic Processes and their Applications, forthcoming.

[49] JACOD, J., AND P. PROTTER (1998). Asymptotic error distribution for the Euler method for stochastic differential equations. Annals of Probability 26, 267–307.

[50] JACOD, J. AND A. N. SHIRYAEV (2003). Limit Theorems for Stochastic Processes. Springer.

[51] KALNINA, I. AND O. B. LINTON (2006). Estimating Quadratic Variation Consistently in the Presence of Correlated Measurement Error. Forthcoming in Journal of Econometrics.

[52] KARATZAS, I. AND S. E. SHREVE (2005). Brownian Motion and Stochastic Calculus, New York: Springer-Verlag.

[53] Kinnebrock, S. and M. Podolskij (2008). An Econometric Analysis of Modulated Realised Covariance, Regression in Noisy Diffusion Models. Working Paper.

[54] KLEIDON, A., AND I. WERNER (1996). U.K. and U.S. trading of British cross-listed stocks: An intraday analysis of market integration. Review of Financial Studies 9, 619-644.

[55] KRISTENSEN, D. (2008). Nonparametric Filtering of the Realised Spot Volatility: A Kernel-Based Approach. Forthcoming in Econometric Theory.

[56] LAHIRI, S. N., M. S. KAISER, N. CRESSIE, AND N. HSU (1999). Prediction of Spatial Cumulative Distribution Functions Using Subsampling. Journal of the American Statistical Association, 94, 86–97

[57] LARGE, G. (2007). Estimating Quadratic Variation When Quoted Prices Change by a Constant Increment. Nuffield College Economics Group Working Paper.

[58] LI, Y., AND P. A. MYKLAND (2006). Determining the Volatility of a Price Process in the presence of rounding errors. Working Paper, Department of Statistics, the University of Chicago.

[59] LI, Y., AND P. A. MYKLAND (2007). Are volatility estimators robust with respect to modelling assumptions? Bernoulli 13 (3), 601-622.

[60] LINTON, O. B., AND I. KALNINA (2007). Discussion of Aït-Sahalia and Shephard. *World Congress of the Econometric Society.* Forthcoming Cambridge University Press

[61] LOCKWOOD, L. J., AND S. C. LINN (1990). An examination of stock market return volatility during overnight and intraday periods, 1964–1989. Journal of Finance 45, 591-601.

[62] MADDALA, G.S. (1977). *Econometrics.* McGraw Hill, New York.

[63] MANCINI, C. (2004). Estimation of the characteristics of the jumps of a general Poisson-diffusion model, Scandinavian Actuarial Journal, 2004, 1: 42-52

[64] MCINISH, T. H., AND R. A. WOOD (1992). An analysis of intraday patterns in bid/ask spreads for NYSE stocks. Journal of Finance 47, 753–764.

[65] MIKOSCH, T. AND C. STARICA (2003). Stock Market Risk-Return Inference. An Unconditional, Non-Parametric Approach. Available at SSRN: http://ssrn.com/abstract=882820

[66] MYKLAND, P., AND L. ZHANG (2006). ANOVA for diffusions and Itô Processes. Annals of Statistics, 34, 1931-1963.

[67] MYKLAND, P. AND L. ZHANG (2007). Locally parametric inference in high frequency data. University of Chicago.

[68] PALANDRI, A. (2006). Consistent Realized Covariance for Asynchronous Observations Contaminated by Market Microstructure Noise. Working Paper

[69] PODOLSKIJ, M. (2006). New Theory on Estimation of Integrated Volatility with Applications. PhD Thesis, Bochum University

[70] PODOLSKIJ, M. AND M. VETTER (2006). Estimation of Volatility Functionals in the Simultaneous Presence of Microstructure Noise and Jumps. Working Paper.

[71] PODOLSKIJ, M. AND D. ZIGGEL (2007). Boostrapping Bipower Variation. Technical Report. Ruhr-University of Bochum.

[72] POLITIS, D. N. AND J. P. ROMANO (1994), "Large sample confidence regions based on subsamples under minimal assumptions." *Annals of Statistics* 22, 2031-2050.

[73] POLITIS, D. N., J. P. ROMANO AND M. WOLF (1999). *Subsampling*, Springer-Verlag, New York.

[74] REVUZ, D., AND M. YOR (2005). *Continuous Martingales and Brownian Motion.* Springer-Verlag, New York.

[75] ROBINSON, P.M. (1986). On the errors in variables problem for time series. Journal of Multivariate Analysis 19, 240–250.

[76] ZHANG, L. (2006). Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. Bernoulli, 12 (6), 1019-1043.

[77] ZHANG, L. (2008). Estimating Covariation: Epps Effect, Microstructure Noise. Working Paper.

[78] ZHANG, L., P. MYKLAND, AND Y. AÏT-SAHALIA (2005a). A tale of two time scales: determining integrated volatility with noisy high-frequency data. Journal of the American Statistical Association, 100, 1394–1411

[79] ZHANG, L., P. MYKLAND, AND Y. AÏT-SAHALIA (2005b). Edgeworth Expansions for Realized Volatility and Related Estimators. Working paper, Princeton University.

[80] ZHOU, B. (1996). High-frequency data and volatility in foreign-exchange rates. Journal of Business and Economic Statistics 14, 45–52.

Appendix A

Appendices for Chapter 1

A.1 Proof of Theorem 1.4.1

We assume for simplicity that $\mu \equiv 0$ in the sequel. Drift is not important in high frequencies as it is of order dt, while the diffusion term is of order $\sqrt{dt}$ (see, for example Aït-Sahalia et al.(2006)). With the assumptions of Theorem, the same method as in the proof can be applied to the drift, yielding the conclusion that it is not important statistically.

Proof of Theorem. We will rely on the first and second moment calculations of $[X,u]^{\{n\}}$, $[u,u]^{\{n\}}$, $[X,u]^{avg}$, $[u,u]^{avg}$, and respective covariances. These can be found in Section A.2. From there, $2\overline{n}^{1/2}[X,u]^{avg} - \frac{\overline{n}^{3/2}}{n}2[X,u]^{\{n\}} = o_p(1)$ by Chebyshev's inequality and similarly $\frac{\overline{n}^{3/2}}{n}[X,X]^{\{n\}} = o_p(1)$. Also, we have $\mathrm{E}\left([X,X]^{avg} - QV_X\right) = o(\overline{n}^{-1/2})$ from ZMA (2005) and $E[\overline{n}^{1/2}[u,u]^{avg} - \frac{\overline{n}^{3/2}}{n}[u,u]^{\{n\}}] = o(1)$. Therefore,

$$
\begin{aligned}
& \overline{n}^{1/2}\left(\widehat{QV}_X - QV_X\right) \\
= \; & \overline{n}^{1/2}\left([X,X]^{avg} - \mathrm{E}\,[X,X]^{avg} + [u,u]^{avg} - E\,[u,u]^{avg} - \frac{1}{K}[u,u]^{\{n\}} + \frac{1}{K}E\,[u,u]^{\{n\}}\right) \\
& + o_p(1).
\end{aligned}
$$

We use Berk's (1973) central limit theorem for m-dependent variables with $m = 1$. Note that we can prove the CLT for the special case $\alpha = 0$ and convergence rate $n^{1/6}$, then get the needed result by multiplying and dividing the main expression by $n^{\alpha/3}$. We proceed in the case where all three terms contribute, which is the case where K is chosen optimally to be $K = O(n^{2/3})$. Also, we can do all calculations, conditional on $\sigma = \{\sigma_t,\, t \in [0,1]\}$. Then, since σ is independent of all other randomness, we can conclude the same CLT unconditionally. We apply Berk's CLT to the following sums of U_{ni},

$$
\begin{aligned}
& T_n \\
= \;& V(\sigma)^{-1}\overline{n}^{1/2}\left([X,X]^{avg} - \mathrm{E}\,[X,X]^{avg} + [u,u]^{avg}\right. \\
& \left. - E\,[u,u]^{avg} - \frac{1}{K}[u,u]^{\{n\}} + \frac{1}{K}E\,[u,u]^{\{n\}}\right) \\
= \;& \overline{n}^{-1/2}\sum_{i=1}^{\overline{n}-1} V(\sigma)^{-1/2}U_{ni},
\end{aligned}
$$

$$
\begin{aligned}
U_{ni} \;=\; & \frac{\overline{n}}{K}\left\{\sum_{j=1}^{K}\left(X_{t_{iK+j}} - X_{t_{(i-1)K+j}}\right)^2 - \int_{t_{(i-1)K+j}}^{t_{iK+j}} \sigma_u^2 du\right\} \\
& + \frac{\overline{n}}{K}\left\{\sum_{j=1}^{K}\left(u_{t_{iK+j}} - u_{t_{(i-1)K+j}}\right)^2 - E\left(u_{t_{iK+j}} - u_{t_{(i-1)K+j}}\right)^2\right\} \\
& - \frac{\overline{n}}{K}\left\{\sum_{s=1}^{2K-1}\frac{1}{2}\left(u_{(i-1)K+1+s} - u_{(i-1)K+s}\right)^2 - E\frac{1}{2}\left(u_{(i-1)K+1+s} - u_{(i-1)K+s}\right)^2\right\} \\
\equiv \;& U_{ni}^{x} + U_{ni}^{u_1} + U_{ni}^{u_2} \equiv U_{ni}^{x} + U_{ni}^{u}.
\end{aligned}
$$

There are 4 conditions to be satisfied in Berk's CLT, which we denote (i)-(iv). Notice that $\{U_{ni}\}_{i=1}^{\overline{n}-1}$ is (conditionally on σ) a sequence of 1-dependent random variables. Therefore, condition (iv) on dependence is trivially satisfied. Condition (iii) requires the following to exist and be non-zero,

$$V(\sigma) = \lim_{n \to \infty} \overline{n}^{-1} \mathrm{var} \left\{ \sum_{i=1}^{\overline{n}-1} U_{ni} \right\}.$$

This follows by our moment calculations,

$$\begin{aligned}
& V(\sigma) \\
= & \lim_{n\to\infty} \left\{ \overline{n} \mathrm{var} [X,X]^{avg} + \overline{n} \mathrm{var} [u,u]^{avg} + \left(\frac{\overline{n}^{3/2}}{n} \right)^2 \mathrm{var} [u,u]^{\{n\}} \right. \\
& \left. -2 \frac{\overline{n}^2}{n} \mathrm{cov} \left([u,u]^{\{n\}}, [u,u]^{avg} \right) \right\} \\
= & \frac{4}{3} \int \sigma_t^4 dt + \frac{2}{c^3} \left(12\delta^4 + 4E\epsilon^4 \int \omega^4 (u) \, du + 24\delta^2 \int \omega^2 (u) \, du \right) \\
& - \frac{2}{c^3} \left(8\delta^4 + 4 \left(E\epsilon^4 - 1 \right) \int \omega^4 (u) \, du + 16\delta^2 \int \omega^2 (u) \, du \right) \\
= & \frac{4}{3} \int \sigma_t^4 dt + c^{-3} \left(8\delta^4 + 16\delta^2 \int \omega^2 (u) \, du + 8 \int \omega^4 (u) \, du \right).
\end{aligned}$$

Condition (ii) requires

$$\mathrm{var} \left(U_{ns+1} + ... + U_{ns'} \right) \leq \left(s' - s \right) M' \text{ for all } i,\ j, \text{ and } n \text{ sufficiently large}, \qquad (A.1)$$

where M' is some constant. We have that

$$\begin{aligned}
& \mathrm{var} \left(U_{ns+1}^x + ... + U_{ns'}^x \right) \\
= & \mathrm{var} \left(\overline{n} \frac{1}{K} \sum_{j=1}^{K} \sum_{i=s+1}^{s'} \left\{ \left(X_{t_{iK+j}} - X_{t_{(i-1)K+j}} \right)^2 - \int_{t_{(i-1)K+j}}^{t_{iK+j}} \sigma_u^2 du \right\} \right) \\
\leq & 2 \left(s' - s \right) \left\{ \sup_{u \in [0,1]} \sigma^2 (u) \right\}
\end{aligned}$$

$$
\begin{aligned}
& \operatorname{var}\left(U_{ns+1}^{u} + \ldots + U_{ns'}^{u}\right) \\
= & \frac{\overline{n}^2}{K^2} \sum_{i=s+1}^{s'} \left\{ 4 \sum_{j=1}^{K} \operatorname{var}\left(u_{iK+j} u_{(i-1)K+j}\right) + \sum_{j=1}^{2K-1} \operatorname{var}\left(u_{(i-1)K+1+j} u_{(i-1)K+j}\right) \right\} \\
& + \frac{\overline{n}^2}{K^2} \sum_{i=s+1}^{s'} \left\{ \frac{1}{4} \operatorname{var}\left(u_{iK-K+1}^2\right) - \frac{3}{4} \operatorname{var}\left(u_{iK+K}^2\right) + \operatorname{var}\left(u_{iK+1}^2\right) \right\} + o(1) \\
\leq & \ (s' - s) C_u \left\{ 6c^{-3} + c^{-4} \right\} + o(1),
\end{aligned}
$$

where the $o(1)$ terms arise from the mean $m(.)$ and are asymptotically negligible, while c is the constant in the definition of K and C_u is the maximum of the upper bound for $(\operatorname{var}(u_i))^2$ and the upper bound for $\operatorname{var}(u_i^2)$. Their respective expressions are as follows:

$$
\begin{aligned}
\operatorname{var}(u_i) & \leq \delta^2 + \left\{ \sup_{t \in [0,1]} \omega(t) \right\}^2 \\
\operatorname{var}\left(u_i^2\right) & \leq 2\delta^4 + 4 \left\{ \sup_{t \in [0,1]} m(t) \right\}^2 \left\{ \sup_{t \in [0,1]} \omega(t) \right\}^2 + 4 \sup_{t \in [0,1]} m(t) \left\{ \sup_{t \in [0,1]} \omega(t) \right\}^3 \mathrm{E}|\epsilon|^3 \\
& + \left\{ \sup_{t \in [0,1]} \omega(t) \right\}^4 (\mathrm{E}\epsilon^4 - 1) + 4 \left\{ \sup_{t \in [0,1]} m(t) \right\}^2 \delta^2 + 4 \left\{ \sup_{t \in [0,1]} \omega(t) \right\}^2 \delta^2.
\end{aligned}
$$

By the Cauchy-Schwarz inequality we obtain (A.1).

Finally, condition (i) is:

$$
\text{For some } \eta > 0 \text{ and } M < \infty, \ \mathrm{E}\left|U_{ni}\right|^{2+\eta} \leq M \text{ for all } i \text{ and } n. \tag{A.2}
$$

Then for some constant C, $E[|U_{ni}|^{2+\eta}] \leq C(E[|U_{ni}^{x}|^{2+\eta}] + E[|U_{ni}^{u_1}|^{2+\eta}] + E[|U_{ni}^{u_2}|^{2+\eta}])$. Different arguments are required for the U_{ni}^{x} and U_{ni}^{u} terms - the summands in U_{ni}^{x} are highly dependent but individually of small order while the summands in U_{ni}^{u} are independent or of low order dependence but of individually larger order. Define

$$
w_{nij} = \frac{n}{K} \left[(X_{t_{iK+j}} - X_{t_{(i-1)K+j}})^2 - \int_{t_{(i-1)K+j}}^{t_{iK+j}} \sigma_u^2 du \right].
$$

Then, since $X_{t_{iK+j}} - X_{t_{(i-1)K+j}} \sim N(0, \int_{t_{(i-1)K+j}}^{t_{iK+j}} \sigma_u^2 du)$, where $\int_{t_{(i-1)K+j}}^{t_{iK+j}} \sigma_u^2 du = O(K/n)$, we have $E[|w_{nij}|^r] \leq C_r < \infty$ for all r, i, j. Note that $X_{t_{iK+j}} - X_{t_{(i-1)K+j}}$ and $X_{t_{iK+j'}} - X_{t_{(i-1)K+j'}}$ for $j \neq j'$ are highly dependent. We write

$$U_{ni}^x = \frac{\overline{n}}{K} \left\{ \sum_{j=1}^{K} \left(X_{t_{iK+j}} - X_{t_{(i-1)K+j}} \right)^2 - \int_{t_{(i-1)K+j}}^{t_{iK+j}} \sigma_u^2 du \right\} = \frac{\overline{n}}{K} \frac{K}{n} \sum_{j=1}^{K} w_{nij}.$$

Therefore, by Minkowski inequality

$$\left(E[|U_{ni}^x|^{2+\eta}] \right)^{1/2+\eta} \leq \frac{\overline{n}K}{Kn} \sum_{j=1}^{K} \left(E\left[|w_{nij}|^{2+\eta} \right] \right)^{1/2+\eta} = \frac{1}{K} \sum_{j=1}^{K} \left(E\left[|w_{nij}|^{2+\eta} \right] \right)^{1/2+\eta} < \infty.$$

Similar arguments apply to the terms $U_{ni}^{u_1}$ and $U_{ni}^{u_2}$, where we make use of the assumption that $E[|\epsilon_{t_i}|^{4+\eta}] < \infty$. We just show the argument for $U_{ni}^{u_1}$. Recall that $u_{t_{iK+j}} - u_{t_{(i-1)K+j}} = v_{t_{iK+j}} - v_{t_{(i-1)K+j}} + \varepsilon_{t_{iK+j}} - \varepsilon_{t_{(i-1)K+j}}$, where $v_{t_i} = \delta\sqrt{n}\left(W_{t_i} - W_{t_{i-1}} \right)$ and $\varepsilon_{t_i} = m(t_i) + \omega(t_i)\,\epsilon_{t_i}$, so it suffices to show this result for the two components. The arguments to do with v_{t_i} are straightforward because all moments exist and the magnitude is just right. Regarding the $\varepsilon_{t_{iK+j}}$ terms, let $\xi_{nj} = (\varepsilon_{t_{iK+j}} - \varepsilon_{t_{(i-1)K+j}})^2 - E(\varepsilon_{t_{iK+j}} - \varepsilon_{t_{(i-1)K+j}})^2$, where ξ_{nj} are independent and mean zero random variables across $j = 1, \ldots, K$. First, notice that $E[|\epsilon_{t_{iK+j}}|^{4+2\eta}] < \infty$, $\sup_{u\in[0,1]} |m(u)| < \infty$, and $\sup_{u\in[0,1]} |\omega(u)| < \infty$ imply that $E[|\varepsilon_{t_{iK+j}}|^{4+2\eta}] < \infty$. Then, $E[|\varepsilon_{t_{iK+j}}|^{4+2\eta}] < \infty$ implies that $E[|(\varepsilon_{t_{iK+j}} - \varepsilon_{t_{(i-1)K+j}})^2|^{2+\eta}] < \infty$ and hence $E[|\xi_{nj}|^{2+\eta}]$. Then, by the Marcinkiewicz-Zygmund inequality for independent random variables and Hölder's inequality for sums

$$\begin{aligned}
E\left[\left|\sum_{j=1}^{K}\xi_{nj}\right|^{p}\right] &\le C_pE\left[\left|\sum_{j=1}^{K}\xi_{nj}^2\right|^{p/2}\right]\\
&\le C_pE\left[\left|\left(\sum_{j=1}^{K}\left|\xi_{nj}\right|^{p}\right)^{2/p}\right|^{p/2}\right]\\
&= C_p\left(\sum_{j=1}^{K}E\left|\xi_{nj}\right|^{p}\right)<\infty
\end{aligned}$$

for any p for which $E\left|\xi_{nj}\right|^{p}<\infty$. It follows that

$$E[|U_{ni}^{\varepsilon}|^{2+\eta}]\equiv E\left[\left|\frac{\overline{n}}{K}\sum_{j=1}^{K}\xi_{nj}\right|^{2+\eta}\right]\le\left(\frac{\overline{n}}{K}\right)^{2+\eta}KC_{2+\eta}<\infty$$

for $K=O(n^{2/3})$. This establishes condition (i).

To conclude, the conditions of Berk's theorem are satisfied conditional on σ and so we have shown that $\Pr(T_n\le t|\sigma)\to\Phi(t)$ for all t, which implies that $\Pr(T_n\le t)\to\Phi(t)$, where $\Phi(t)$ denotes the c.d.f. of a standard normal random variable.

A.2 Technical Appendix to Chapter 1

This appendix contains first and second moment calculations of components of the estimator of Chapter 1.

For deterministic sequences A_n, B_n we use the notation $A_n \sim B_n$ to mean that A_n is equal to B_n plus something of smaller order than B_n, i.e., $A_n/B_n\to 1$.

To follow easier the notation regarding all subscripts, it is convenient to think in terms of grids. The time indices of the full dataset with n data points are on a grid $\mathcal{G}=$

$\{1,2,3,\ldots,n\}$. For the first few lemmas we take the first subsample only, which has time indices on the first subgrid $\mathcal{G}_1 = \{1, K+1, 2K+1, \ldots, (\overline{n}-1)K+1\}$, where $\overline{n} = n/K$. This translates into $(i-1)K+1$, $i = 1,\ldots,\overline{n}$. Hence, for summations like the one defining $[Y,Y]^{n_1}$ we will need to take

$$\left\{\begin{array}{c} K+1, 2K+1, \ldots, (\overline{n}-1)K+1 \\ 1, K+1, \ldots, (\overline{n}-2)K+1 \end{array}\right\} = \left\{\begin{array}{c} iK+1, i=1,\ldots,\overline{n}-1 \\ (i-1)K+1, i=1,\ldots,\overline{n}-1 \end{array}\right\}.$$

Similarly, the j^{th} subgrid is $\mathcal{G}_j = \{j, K+j, 2K+j, \ldots, (\overline{n}-1)K+j\}, j = 1,2,\ldots,K$. This translates into $(i-1)K+j$, $i = 1,\ldots,\overline{n}$. Hence,

$$\left\{\begin{array}{c} K+j, 2K+j, \ldots, (\overline{n}-1)K+j \\ j, K+j, \ldots, (\overline{n}-2)K+j \end{array}\right\} = \left\{\begin{array}{c} iK+j, i=1,\ldots,\overline{n}-1 \\ (i-1)K+j, i=1,\ldots,\overline{n}-1 \end{array}\right\}.$$

We assume for simplicity that $\mu \equiv 0$ in the sequel. Drift is not important in high frequencies as it is of order dt, while the diffusion term is of order $\sqrt{dt}$ (see, for example Aït-Sahalia et al.(2006)).With the assumptions of Theorem 1.4.1, the same method as in the proof can be applied to the drift, yielding the conclusion that it is not important statistically.

PROOF OF THEOREM 1.4.1. Expectations are taken conditional on the whole path of σ_t. We have

$$\begin{aligned}
\overline{n}^{1/2}\left(\widehat{QV}_X - QV_X\right) &= \overline{n}^{1/2}[Y,Y]^{avg} - \frac{\overline{n}^{3/2}}{n}[Y,Y]^n - \overline{n}^{1/2}QV_X \\
&= \overline{n}^{1/2}\{[X,X]^{avg} + 2[X,u]^{avg} + [u,u]^{avg}\} \\
&\quad -\frac{\overline{n}^{3/2}}{n}\left\{[X,X]^{\{n\}} + 2[X,u]^{\{n\}} + [u,u]^{\{n\}}\right\} - \overline{n}^{1/2}QV_X \\
&\equiv C1 + C2 + C3 - C4 - C5 - C6.
\end{aligned}$$

We calculate the order in probability of these terms by computing their means and variances and using Chebychev's inequality. We show below that:

		mean	variance
C1	$\overline{n}^{1/2}[X,X]^{avg} - \overline{n}^{1/2}QV_X$	0	$O(1)$
C2	$2\overline{n}^{1/2}[X,u]^{avg}$	$O(n^{\alpha/2})$	$o(1)$
C3	$\overline{n}^{1/2}[u,u]^{avg}$	$O(n^{1/2})$	$O(1)$
C4	$\frac{\overline{n}^{3/2}}{n}[X,X]^{\{n\}}$	$o(1)$	$o(1)$
C5	$\frac{\overline{n}^{3/2}}{n}2[X,u]^{\{n\}}$	$O(n^{\alpha/2})$	$o(1)$
C6	$\frac{\overline{n}^{3/2}}{n}[u,u]^{\{n\}}$	$O(n^{1/2})$	$O(1)$.

(C1) The term $\overline{n}^{1/2}[X,X]^{\overline{n}}_{avg} - \overline{n}^{1/2}QV_X$ has zero mean and variance $O(1)$ from the result in Zhang et al. (2005) (eqn. 49, pp. 1401), and:

$$\overline{n}^{1/2}\left([X,X]^{\overline{n}}_{avg} - QV_X\right) \Longrightarrow N\left(0, \frac{4}{3}\int_0^1 \sigma_t^4 dt\right).$$

(C2,C5) In Lemma A5 we show that $E[C2 - C5] = o_p(1)$. Lemma A2 shows that variance of $C5$ is small. From Lemma A6, $4\overline{n}\text{var}[X,u]^{avg} = O\left(\overline{n}n^{\beta-2}\right) = O\left(n^{-1}\right)$.

(C3,C6) In Lemma A7 we show that $E[C3 - C6] = o_p(1)$. From Lemma A4, $\text{var}\left(\frac{\overline{n}^{3/2}}{n}[u,u]^{\{n\}}\right) = O(1)$ and from Lemma A8, $\text{var}\left(\overline{n}^{1/2}[u,u]^{avg}\right) = O(1)$.

(C4) By Jacod and Protter (1998), $[X,X]^n - QV_X = O_p\left(n^{-1/2}\right)$ and so $\frac{\overline{n}^{3/2}}{n}[X,X]^n = O_p\left(\frac{\overline{n}^{3/2}}{n}\right) = O_p\left(n^{\alpha-1/2}\right) = o_p(1)$. Since $[X,X]^n - [X,X]^{\{n\}} = O_p(K/n)$, we have also $\frac{\overline{n}^{3/2}}{n}[X,X]^{\{n\}} = o_p(1)$.

It follows that the limiting distribution of $\overline{n}^{1/2}(\widehat{QV}_X - QV_X)$ is that of $C1+C3-C6$. The covariances between these terms are calculated in Lemmas B1, B2, and B3.

Therefore, the asymptotic variance is

$$
\begin{aligned}
V &= \lim_{n\to\infty}\left\{\overline{n}\mathrm{var}\left[X,X\right]^{avg}+\overline{n}\mathrm{var}\left[u,u\right]^{avg}+\left(\frac{\overline{n}^{3/2}}{n}\right)^{2}\mathrm{var}\left[u,u\right]^{\{n\}}-2\frac{\overline{n}^{2}}{n}\mathrm{cov}\left(\left[u,u\right]^{\{n\}},\left[u,u\right]^{avg}\right)\right. \\
&= \frac{4}{3}\int_0^1\sigma_t^4 dt+2c^{-3}\left(12\delta^4+4E\epsilon^4\int\omega^4\left(u\right)du+24\delta^2\int\omega^2\left(u\right)du\right) \\
&\quad -2c^{-3}\left(8\delta^4+4\left(E\epsilon^4-1\right)\int\omega^4\left(u\right)du+16\delta^2\int\omega^2\left(u\right)du\right) \\
&= \frac{4}{3}\int_0^1\sigma_t^4 dt+c^{-3}\left(8\delta^4+16\delta^2\int\omega^2\left(u\right)du+8\int\omega^4\left(u\right)du\right).
\end{aligned}
$$

A.2.1 Lemmas

Here we give the lemmas needed in the proof of the Theorem 1.4.1.

LEMMA A1. *For all n*

$$E[X,u]^{n_1} = \delta\gamma_n \sum_{i=1}^{n_1-1} \int_{t_{iK}}^{t_{iK+1}} \sigma_t dt.$$

LEMMA A2. *As $n \to \infty$,*

$$\begin{aligned} \operatorname{var}[X,u]^{n_1} &= O\left(n^{-\alpha}\right) + O\left(\frac{1}{n_1^2}\right) \\ 4\left(\frac{\overline{n}^{3/2}}{n}\right)^2 \operatorname{var}\left[X,u\right]^{\{n\}} &= o\left(1\right). \end{aligned}$$

LEMMA A3. *As $n \to \infty$,*

$$E[u,u]^{n_1} = [m,m]^{n_1} + 2n_1 n^{-\alpha} \int_0^1 \omega^2\left(u\right) du + 2n_1\delta^2 n^{-\alpha} + O\left(n^{-\alpha}\right).$$

LEMMA A4. *As $n \to \infty$,*

$$\begin{aligned} &\operatorname{var}[u,u]^{n_1} \\ = \ &\frac{n_1}{n^{2\alpha}}\left\{12\delta^4 + 4E\epsilon^4 \int \omega^4\left(u\right) du + 24\delta^2 \int \omega^2\left(u\right) du + O\left(\frac{n^{\alpha/2}}{n_1}\right) + O\left(\frac{n^\alpha}{n_1^2}\right)\right\}, \\ &\left(\frac{\overline{n}^{3/2}}{n}\right)^2 \operatorname{var}[u,u]^{\{n\}} \\ = \ &c^{-3}\left\{12\delta^4 + 4E\epsilon^4 \int \omega^4\left(u\right) du + 24\delta^2 \int \omega^2\left(u\right) du\right\} + o\left(1\right) \end{aligned}$$

LEMMA A5. *As* $n \to \infty$,

$$E[X,u]^{avg} = \frac{\delta \gamma_n}{K} \int_0^1 \sigma_t dt + O\left(n^{-\frac{1+\alpha}{2}}\right)$$
$$\overline{n}^{1/2} E[X,u]^{avg} = \frac{\overline{n}^{3/2}}{n} E[X,u]^{\{n\}} + o(1)$$

LEMMA A6. *As* $n \to \infty$,

$$\mathrm{var}[X,u]^{avg} \sim \frac{1}{K} \mathrm{var}[X,u]^{n_1} = O\left(n^{\beta-2}\right).$$

LEMMA A7. *As* $n \to \infty$,

$$\overline{n}^{1/2} E\left[u,u\right]^{avg} = \frac{\overline{n}^{3/2}}{n} E\left[u,u\right]^{\{n\}} + o(1)$$

LEMMA A8. *As* $n \to \infty$,

$$\begin{aligned} \mathrm{var}[u,u]^{avg} &= \frac{n_1}{Kn^{2\alpha}} \left\{ 12\delta^4 + 4E\epsilon^4 \int \omega^4(u)\,du + 24\delta^2 \int \omega^2(u)\,du + O\left(\frac{n^{\alpha/2}}{n_1}\right) + O\left(\frac{n^{\alpha}}{n_1^2}\right) \right\} \\ &= c^{-3}\overline{n}^{-1} \left\{ 12\delta^4 + 4E\epsilon^4 \int \omega^4(u)\,du + 24\delta^2 \int \omega^2(u)\,du + o(1) \right\} \text{ if } \beta = \frac{2}{3}(1-\alpha). \end{aligned}$$

LEMMA B1. *As* $n \to \infty$,

$$\mathrm{cov}\left([X,X]^{avg}, [u,u]^{avg}\right) = O\left(n^{-\alpha}K^{-1}\right) = o\left(\overline{n}^{-1/2}\right).$$

LEMMA B2. *As* $n \to \infty$,

$$\mathrm{cov}\left([X,X]^{avg}, [u,u]^{\{n\}}\right) = O\left(n^{-\alpha}K^{-1}\right) = o\left(\overline{n}^{-1/2}\right).$$

LEMMA B3. *As* $n \to \infty$,

$$\text{cov}\left([u,u]^{\{n\}},[u,u]^{avg}\right)=\frac{n}{\overline{n}^2}c^{-3}\left\{8\delta^4+4\left(E\epsilon^4-1\right)\int\omega^4(u)\,du+16\delta^2\int\omega^2(u)\,du+o(1)\right\}.$$

A.2.2 Proofs of Lemmas

Write symbolically $[X,u]=[X,v]+[X,\varepsilon]$ and $[u,u]=[v,v]+[\varepsilon,\varepsilon]+2[v,\varepsilon]$, where the process X is independent of the process ε and the process v is also independent of the process ε. Also use for a function g and lag $J=1,\ldots,K$, $\Delta_J g(t_i)=g(t_i)-g(t_{i-J})$ with $\Delta=\Delta_1$ for simplicity.

PROOF OF LEMMA A1. We have

$$\begin{aligned}
&E[X,u]^{n_1}\\
=&\sum_{i=1}^{n_1-1}E\left[\left(u_{t_{iK+1}}-u_{t_{(i-1)K+1}}\right)\left(X_{t_{iK+1}}-X_{t_{(i-1)K+1}}\right)\right]\\
=&\sum_{i=1}^{n_1-1}E\left[\int_{t_{(i-1)K+1}}^{t_{iK+1}}\sigma_t dW_t\left[\delta\gamma_n\left(W_{t_{iK+1}}-W_{t_{iK}}\right)\right.\right.\\
&\left.\left.-\delta\gamma_n\left(W_{t_{(i-1)K+1}}-W_{t_{(i-1)K}}\right)+\left(\varepsilon_{t_{iK+1}}-\varepsilon_{t_{(i-1)K+1}}\right)\right]\right]\\
=&\delta\gamma_n\sum_{i=1}^{n_1-1}\int_{t_{iK}}^{t_{iK+1}}\sigma_t dt.
\end{aligned}$$

PROOF OF LEMMA A2. We have

$$
\begin{aligned}
\text{var}[X,u]^{n_1} &= \text{var}[X,v]^{n_1} + \text{var}[X,\varepsilon]^{n_1} + 2\text{cov}\left([X,v]^{n_1},[X,\varepsilon]^{n_1}\right) \\
&= \text{var}[X,v]^{n_1} + \text{var}[X,\varepsilon]^{n_1} \text{ since } \mathrm{E}\left((\Delta X)^2 \Delta W\right) = 0 \text{ by normality} \\
&= O\left(n^{-\alpha}\right) + O\left(\frac{1}{\overline{n}^2}\right). \qquad \text{(A.3)}
\end{aligned}
$$

We prove (A.3) below. First part is $\text{var}\,[X,v]^{n_1} \sim n^{1-\alpha}\left(\frac{1}{n}+\frac{n_1}{n^2}\right) \sim n^{-\alpha}$ by

$$
\begin{aligned}
& \frac{1}{\delta^2\gamma_n^2}\text{var}\,[X,v]^{n_1} \\
= & \frac{1}{\delta^2\gamma_n^2}\text{var}\left[\sum_{i=1}^{n_1-1}\left(X_{t_{iK+1}} - X_{t_{(i-1)K+1}}\right)\left(v_{t_{iK+1}} - v_{t_{(i-1)K+1}}\right)\right] \\
= & \text{var}\left[\sum_{i=1}^{n_1-1}\left(X_{t_{iK+1}} - X_{t_{(i-1)K+1}}\right)\left(\left\{W_{t_{iK+1}} - W_{t_{iK}}\right\} - \left\{W_{t_{(i-1)K+1}} - W_{t_{(i-1)K}}\right\}\right)\right] \\
= & \sum_{i=1}^{n_1-1}\text{var}\left[\left(X_{t_{iK+1}} - X_{t_{(i-1)K+1}}\right)\left(W_{t_{iK+1}} - W_{t_{iK}}\right)\right] \\
& + \sum_{i=1}^{n_1-1}\text{var}\left[\left(X_{t_{iK+1}} - X_{t_{(i-1)K+1}}\right)\left(W_{t_{(i-1)K+1}} - W_{t_{(i-1)K}}\right)\right] \\
& + 2\sum_{i=1}^{n_1-2}\mathrm{E}\left[\Delta_K X_{t_{iK+1}}\left(\Delta W_{t_{iK+1}} - \Delta W_{t_{(i-1)K+1}}\right)\Delta_K X_{t_{(i+1)K+1}}\left(\Delta W_{t_{(i+1)K+1}} - \Delta W_{t_{iK+1}}\right)\right] \\
= & \frac{2}{n}\int_{t_1}^{t_{n-K+1}}\sigma_t^2 dt + O\left(\frac{n_1}{n^2}\right),
\end{aligned}
$$

where for the final equality we use:

$$
\begin{aligned}
&\sum_{i=1}^{n_1-1} \operatorname{var}\left[\left(X_{t_{iK+1}} - X_{t_{(i-1)K+1}}\right)\left(W_{t_{iK+1}} - W_{t_{iK}}\right)\right] \\
= &\sum_{i=1}^{n_1-1} E\left(X_{t_{iK+1}} - X_{t_{(i-1)K+1}}\right)^2 E\left(W_{t_{iK+1}} - W_{t_{iK}}\right)^2 \\
&+ \sum_{i=1}^{n_1-1} E^2\left(X_{t_{iK+1}} - X_{t_{(i-1)K+1}}\right)\left(W_{t_{iK+1}} - W_{t_{iK}}\right) \text{ by normality} \\
= &\sum_{i=1}^{n_1-1}\left[\int_{t_{(i-1)K+1}}^{t_{iK+1}} \sigma_t^2 dt \frac{1}{n} + \left(\int_{t_{iK}}^{t_{iK+1}} \sigma_t dt\right)^2\right] = \frac{1}{n}\int_{t_1}^{t_{n-K+1}} \sigma_t^2 dt + O\left(\frac{n_1}{n^2}\right),
\end{aligned}
$$

$$
\begin{aligned}
&\sum_{i=1}^{n_1-1} \operatorname{var}\left[\left(X_{t_{iK+1}} - X_{t_{(i-1)K+1}}\right)\left(W_{t_{(i-1)K+1}} - W_{t_{(i-1)K}}\right)\right] \\
= &\sum_{i=1}^{n_1-1} \int_{t_{(i-1)K+1}}^{t_{iK+1}} \sigma_t^2 dt \frac{1}{n} = \frac{1}{n}\int_{t_1}^{t_{n-K+1}} \sigma_t^2 dt,
\end{aligned}
$$

and

$$
\begin{aligned}
&2\sum_{i=1}^{n_1-2} \mathrm{E}\left[\Delta_K X_{t_{iK+1}}\left(\Delta W_{t_{iK+1}} - \Delta W_{t_{(i-1)K+1}}\right)\Delta_K X_{t_{(i+1)K+1}}\left(\Delta W_{t_{(i+1)K+1}} - \Delta W_{t_{iK+1}}\right)\right] \\
= &\ 2\sum_{i=1}^{n_1-2} \mathrm{E}\left(\Delta_K X_{t_{iK+1}} \Delta W_{t_{iK+1}}\right) \mathrm{E}\left(\Delta_K X_{t_{(i+1)K+1}} \Delta W_{t_{(i+1)K+1}}\right) \\
= &\ 2\sum_{i=1}^{n_1-2}\left[\int_{t_{iK}}^{t_{iK+1}} \sigma_t dt \int_{t_{(i+1)K}}^{t_{(i+1)K+1}} \sigma_t dt\right] = O\left(\frac{n_1}{n^2}\right).
\end{aligned}
$$

For the second part of (A.3), we have $\text{var}[X,\varepsilon]^{n_1} = O\left(\frac{1}{\overline{n}^2}\right) + O\left(n^{-\alpha}\right)$ by

$$
\begin{aligned}
& \text{var}[X,\varepsilon]^{n_1} \\
= \; & \text{var}\left[\sum_{i=1}^{n_1-1}\left(X_{t_{iK+1}} - X_{t_{(i-1)K+1}}\right)\left(\varepsilon_{t_{iK+1}} - \varepsilon_{t_{(i-1)K+1}}\right)\right] \\
= \; & \sum_{i=1}^{n_1-1} \text{var}\left[\Delta_K X_{t_{iK+1}} \Delta_K \varepsilon_{t_{iK+1}}\right] + 2\sum_{i=2}^{n_1-1} \text{cov}\left[\Delta_K X_{t_{iK+1}} \Delta_K \varepsilon_{t_{iK+1}}, \Delta_K X_{t_{(i-1)K+1}} \Delta_K \varepsilon_{t_{(i-1)K+1}}\right] \\
= \; & \sum_{i=1}^{n_1-1} \mathrm{E}\left(\Delta_K X_{t_{iK+1}}\right)^2 \mathrm{E}\left(\Delta_K \varepsilon_{t_{iK+1}}\right)^2 \\
& +2\sum_{i=2}^{n_1-1} \mathrm{E}\left(\Delta_K X_{t_{iK+1}}\right) \mathrm{E}\left(\Delta_K X_{t_{(i-1)K+1}}\right) \mathrm{E}\left(\Delta_K \varepsilon_{t_{iK+1}} \Delta_K \varepsilon_{t_{(i-1)K+1}}\right) \\
= \; & \sum_{i=1}^{n_1-1} \int_{t_{(i-1)K+1}}^{t_{iK+1}} \sigma_t^2 dt \left[\left(\Delta_K m_{t_{iK+1}}\right)^2 + n^{-\alpha}\left(\omega^2_{t_{iK+1}} + \omega^2_{t_{(i-1)K+1}}\right)\right] \\
& +2\sum_{i=2}^{n_1-1} \int_{t_{(i-1)K+1}}^{t_{iK+1}} \sigma_t dt \int_{t_{(i-1)K+1}}^{t_{iK+1}} \sigma_t dt \left[\Delta_K m_{t_{iK+1}} \Delta_K m_{t_{(i-1)K+1}} + n^{-\alpha}\omega^2_{t_{(i-1)K+1}}\right] \\
= \; & O\left(\frac{1}{\overline{n}^2}\right) + O\left(n^{-\alpha}\right).
\end{aligned}
$$

Now we prove the second part of Lemma A2. Note that by substituting n for n_1 we get $\text{var}[X,u]^n = O\left(n^{-\alpha}\right) + O\left(\frac{1}{n^2}\right)$, and so $\left(\frac{\overline{n}^{3/2}}{n}\right)^2 \text{var}[X,u]^n \sim (\overline{n}^3/n^2 n^\alpha) + (\overline{n}^3/n^5) \sim n^{3(1-\beta)-2-\alpha} = n^{1-\alpha-3\beta} = n^{1-\alpha-2(1-\alpha)} = n^{\alpha-1} = o(1)$. Since $[X,u]^n - [X,u]^{\{n\}}$ is of smaller order than $[X,u]^{\{n\}}$, the same holds for $\left(\frac{\overline{n}^{3/2}}{n}\right)^2 \text{var}[X,u]^{\{n\}}$.

PROOF OF LEMMA A3. We have

$$
\begin{aligned}
\mathrm{E}[u,u]^{n_1} &= \sum_{i=1}^{n_1-1} \mathrm{E}\left[\left(u_{t_{iK+1}} - u_{t_{(i-1)K+1}}\right)^2\right] \\
&= \sum_{i=1}^{n_1-1} \mathrm{E}\left[\delta^2\gamma_n^2 \left(W_{t_{iK+1}} - W_{t_{iK}}\right)^2 + \right. \\
&\quad \left. \delta^2\gamma_n^2 \left(W_{t_{(i-1)K+1}} - W_{t_{(i-1)K}}\right)^2 + \left(\varepsilon_{t_{iK+1}} - \varepsilon_{t_{(i-1)K+1}}\right)^2\right] \qquad \text{(A.4)} \\
&= \sum_{i=1}^{n_1-1} \mathrm{E}\left[\left(\varepsilon_{t_{iK+1}} - \varepsilon_{t_{(i-1)K+1}}\right)^2\right] + \frac{2}{n}\left(n_1 - 1\right)\delta^2\gamma_n^2 \\
&= [m,m]^{n_1} + 0 \qquad \text{(A.5)} \\
&\quad + n^{-\alpha} \sum_{i=1}^{n_1-1} \left[\omega^2\left(\frac{iK+1}{n}\right) + \omega^2\left(\frac{(i-1)K+1}{n}\right)\right] + \frac{2}{n}\left(n_1 - 1\right)\delta^2\gamma_n^2 \qquad \text{(A.6)} \\
&= [m,m]^{n_1} + 2n_1 n^{-\alpha} \int_0^1 \omega^2(u)\,du + O\left(n^{-\alpha}\right) + 2n_1\delta^2 n^{-\alpha}.
\end{aligned}
$$

We prove (A.5) below.

$$
\begin{aligned}
& \sum_{i=1}^{n_1-1} E\left[\left(\varepsilon_{t_{iK+1}} - \varepsilon_{t_{(i-1)K+1}}\right)^2\right] \\
= & \sum_{i=1}^{n_1-1} E\left[\left(m_{t_{iK+1}} - m_{t_{(i-1)K+1}} + n^{-\alpha/2}\omega\left(\frac{iK+1}{n}\right)\epsilon_{t_{iK+1}} - n^{-\alpha/2}\omega\left(\frac{(i-1)K+1}{n}\right)\epsilon_{t_{(i-1)K+1}}\right)^2\right] \\
= & \left(m_{t_{iK+1}} - m_{t_{(i-1)K+1}}\right)^2 + n^{-\alpha} \sum_{i=1}^{n_1-1} \left[\omega^2\left(\frac{iK+1}{n}\right) + \omega^2\left(\frac{(i-1)K+1}{n}\right)\right].
\end{aligned}
$$

PROOF OF LEMMA A4. We have

$$
\begin{aligned}
& \operatorname{var}[u,u]^{n_1} \\
= \; & \operatorname{var}[v,v]^{n_1} + \operatorname{var}[e,e]^{n_1} + 4\operatorname{var}[v,e]^{n_1} \\
= \; & 12\delta^4 n^{-2\alpha} n_1 + O\left(n^{-2\alpha}\right) & \text{(A.7)} \\
& +4n_1 n^{-2\alpha} E\epsilon^4 \textstyle\int \omega^4(u)\,du + O\left(n^{-3\alpha/2}\right) + O\left(n^{-\alpha}\overline{n}^{-1}\right) \\
& +24\delta^2 n^{-2\alpha} n_1 \textstyle\int \omega^2(u)\,du + O\left(n^{-2\alpha}\right) + O\left(n^{-\alpha}\overline{n}^{-1}\right) \\
= \; & 12\delta^4 n^{-2\alpha} n_1 + 24\delta^2 n^{-2\alpha} n_1 \textstyle\int \omega^2(u)\,du \\
& +4n_1 n^{-2\alpha} E\epsilon^4 \textstyle\int \omega^4(u)\,du + O\left(n^{-3\alpha/2}\right) + O\left(n^{-\alpha}\overline{n}^{-1}\right) & \text{(A.8)} \\
= \; & \frac{n_1}{n^{2\alpha}} \left\{ 12\delta^4 + 4E\epsilon^4 \textstyle\int \omega^4(u)\,du + 24\delta^2 \textstyle\int \omega^2(u)\,du + O\left(\frac{n^{\alpha/2}}{n_1}\right) + O\left(\frac{n^{\alpha}}{n_1^2}\right) \right\}.
\end{aligned}
$$

We prove (A.7) below in a series of steps, but first we derive the second result of Lemma A4. Note that note by substituting n for n_1 we get the following expression for $\operatorname{var}[u,u]^n$,

$$
\operatorname{var}[u,u]^n = \frac{n}{n^{2\alpha}} \left\{ 12\delta^4 + 4E\epsilon^4 \int \omega^4(u)\,du + 24\delta^2 \int \omega^2(u)\,du + O\left(\frac{n^{\alpha/2}}{n}\right) + O\left(\frac{n^{\alpha}}{n^2}\right) \right\}.
$$

From this, we have

$$
\begin{aligned}
& \left(\frac{\overline{n}^{3/2}}{n}\right)^2 \operatorname{var}[u,u]^n \\
= \; & \left(\frac{\overline{n}^{3/2}}{n}\right)^2 \frac{n}{n^{2\alpha}} \left\{ 12\delta^4 + 4E\epsilon^4 \textstyle\int \omega^4(u)\,du + 24\delta^2 \textstyle\int \omega^2(u)\,du + O\left(\frac{n^{\alpha/2}}{n}\right) + O\left(\frac{n^{\alpha}}{n^2}\right) \right\} \\
= \; & \frac{\overline{n}^3}{n n^{2\alpha}} \left\{ 12\delta^4 + 4E\epsilon^4 \textstyle\int \omega^4(u)\,du + 24\delta^2 \textstyle\int \omega^2(u)\,du + o(1) \right\} \\
= \; & c^{-3} \left\{ 12\delta^4 + 4E\epsilon^4 \textstyle\int \omega^4(u)\,du + 24\delta^2 \textstyle\int \omega^2(u)\,du \right\} + o(1),
\end{aligned}
$$

since $\frac{\overline{n}^3}{nn^{2\alpha}} = c^{-3} n^{3(1-\beta)-1-2\alpha} = c^{-3} n^{2-2\alpha-2\frac{3}{3}(1-\alpha)} = c^{-3}$. We get the second result of Lemma A4 by noting that $[u,u]^n - [u,u]^{\{n\}}$ is of smaller order than $[u,u]^{\{n\}}$, so

$\left(\frac{\overline{n}^{3/2}}{n}\right)^2 \text{var}[u,u]^n$ has the same leading term as $\left(\frac{\overline{n}^{3/2}}{n}\right)^2 \text{var}[u,u]^{\{n\}}$.

Now we prove (A.7) by calculating separately each of the three components of $\text{var}[u,u]^{n_1}$.

The first component of $[u,u]^{n_1}$ is

$$\begin{aligned}
[v,v]^{n_1} &= \sum_{i=1}^{n_1-1}\left(v_{t_{iK+1}} - v_{t_{(i-1)K+1}}\right)^2 \\
&= \delta^2\gamma_n^2 \sum_{i=1}^{n_1-1}\left(\left(W_{t_{iK+1}} - W_{t_{iK}}\right) - \left(W_{t_{(i-1)K+1}} - W_{t_{(i-1)K}}\right)\right)^2 \\
&\stackrel{d}{=} \delta^2\gamma_n^2 \sum_{i=1}^{n_1-1} Z_i^2, \text{ where } Z_i \text{ are 1-dependent } N\left(0,\frac{2}{n}\right), \text{cov}(Z_i^2, Z_{i-1}^2) = \frac{2}{n^2},
\end{aligned}$$

and hence,

$$\begin{aligned}
\text{var}[v,v]^{n_1} &= \delta^4\gamma_n^4\left(n_1-1\right)\times 2\left(\frac{2}{n}\right)^2 + 2\delta^4\gamma_n^4\left(n_1-2\right)\frac{2}{n^2} \\
&= 12\delta^4\gamma_n^4\left(n_1-1\right)\frac{1}{n^2} + O\left(n^{-2\alpha}\right) \\
&= 12\delta^4 n^{-2\alpha} n_1 + O\left(n^{-2\alpha}\right).
\end{aligned}$$

The second component of $\text{var}[u,u]^{n_1}$ is

$$\begin{aligned}
\text{var}[e,e]^{n_1} &= \text{var}\sum_{i=1}^{n_1-1}\left(\varepsilon_{t_{iK+1}} - \varepsilon_{t_{(i-1)K+1}}\right)^2 \\
&= \sum_{i=1}^{n_1-1}\text{var}\left(\varepsilon_{t_{iK+1}} - \varepsilon_{t_{(i-1)K+1}}\right)^2 + 2\sum_{i=2}^{n_1-1}\text{cov}_{i,i-1} \qquad \text{(A.9)} \\
&= \sum_{i=1}^{n_1-1}\left\{2\omega_{t_{iK+1}}^4 n^{-2\alpha}\left(E\epsilon^4+1\right) + O\left(n^{-\alpha}\frac{K^2}{n^2}\right) + O\left(n^{-\frac{3}{2}\alpha}\frac{K}{n}\right)E\epsilon^3\right\} \\
&\quad +2\sum_{i=2}^{n_1-1} n^{-2\alpha}\omega_{t_{iK+1}}^4\left(E\epsilon^4-1\right) + O\left(n^{-3\alpha/2}\right) + O\left(n^{-\alpha}\overline{n}^{-1}\right) \\
&= 4n^{-2\alpha}E\epsilon^4\sum_{i=1}^{n_1}\omega_{t_{iK+1}}^4 + O\left(n^{-3\alpha/2}\right) + O\left(n^{-\alpha}\overline{n}^{-1}\right) \\
&= 4n_1 n^{-2\alpha}E\epsilon^4\int\omega^4(u)\,du + O\left(n^{-3\alpha/2}\right) + O\left(n^{-\alpha}\overline{n}^{-1}\right),
\end{aligned}$$

where we denote by $\text{cov}_{i,i-1}$ the terms that appear because the sum in $var[e,e]^{n_1}$ involves 1-dependent terms. For exact expression and calculation of the second term in (A.9) see below. Before that, the first term in (A.9) is

$$
\begin{aligned}
& \text{var}\left(\varepsilon_{t_{iK+1}} - \varepsilon_{t_{(i-1)K+1}}\right)^2 \\
= \; & \text{var}\left(\varepsilon_i - \varepsilon_{i-1}\right)^2 \text{ by an obvious change of notation} \\
= \; & \text{var}\left(m_i - m_{i-1} + n^{-\alpha/2}\left(\omega_i \epsilon_i - \omega_{i-1}\epsilon_{i-1}\right)\right)^2 \\
= \; & \text{var}\left(2n^{-\alpha/2}\left(m_i - m_{i-1}\right)\left(\omega_i \epsilon_i - \omega_{i-1}\epsilon_{i-1}\right) + n^{-\alpha}\left(\omega_i \epsilon_i - \omega_{i-1}\epsilon_{i-1}\right)^2\right) \\
= \; & 4n^{-\alpha}\left(m_i - m_{i-1}\right)^2 \text{var}\left(\omega_i \epsilon_i - \omega_{i-1}\epsilon_{i-1}\right) + n^{-2\alpha}\text{var}\left(\omega_i \epsilon_i - \omega_{i-1}\epsilon_{i-1}\right)^2 + \\
& +4n^{-\frac{3}{2}\alpha}\left(m_i - m_{i-1}\right) \text{cov}\left(\omega_i \epsilon_i - \omega_{i-1}\epsilon_{i-1}, \left(\omega_i \epsilon_i - \omega_{i-1}\epsilon_{i-1}\right)^2\right) \\
= \; & 4n^{-\alpha}\left(m_i - m_{i-1}\right)^2\left(\omega_i^2 + \omega_{i-1}^2\right) + n^{-2\alpha}\left\{\left(\omega_i^4 + \omega_{i-1}^4\right)\left(E\epsilon^4 - 1\right) + 4\omega_i^2\omega_{i-1}^2\right\} + \\
& +4n^{-\frac{3}{2}\alpha}\left(m_i - m_{i-1}\right)\left(\omega_i^3 + \omega_{i-1}^3\right) E\epsilon^3 \\
= \; & O\left(n^{-\alpha}\frac{K^2}{n^2}\right) + n^{-2\alpha}\left\{\left(\omega_i^4 + \omega_{i-1}^4\right)\left(E\epsilon^4 - 1\right) + 4\omega_i^2\omega_{i-1}^2\right\} + O\left(n^{-\frac{3}{2}\alpha}\frac{K}{n}\right) E\epsilon^3 \\
= \; & 2\omega_{t_{iK+1}}^4 n^{-2\alpha}\left(E\epsilon^4 + 1\right) + O\left(n^{-\alpha}\frac{K^2}{n^2}\right) + O\left(n^{-\frac{3}{2}\alpha}\frac{K}{n}\right) E\epsilon^3.
\end{aligned}
$$

The second term in (A.9) is

$$\begin{aligned}
\mathrm{cov}_{i,i-1} &= \mathrm{cov}\left[\left(\varepsilon_{t_{iK+1}} - \varepsilon_{t_{(i-1)K+1}}\right)^2, \left(\varepsilon_{t_{(i-1)K+1}} - \varepsilon_{t_{(i-2)K+1}}\right)^2\right] \\
&= \mathrm{cov}\left[(\varepsilon_3 - \varepsilon_2)^2, (\varepsilon_2 - \varepsilon_1)^2\right] \text{ (change of notation)} \\
&= \mathrm{var}\left(\varepsilon_2^2\right) - 2\mathrm{cov}\left(\varepsilon_2^2, \varepsilon_1\varepsilon_2\right) - 2\mathrm{cov}\left(\varepsilon_2\varepsilon_3, \varepsilon_2^2\right) + 4\mathrm{cov}\left(\varepsilon_1\varepsilon_2, \varepsilon_2\varepsilon_3\right) \\
&= 4m_2^2 n^{-\alpha}\omega_2^2 + n^{-2\alpha}\omega_2^4\left(E\epsilon^4 - 1\right) + 4m_2 n^{-3\alpha/2}\omega_2^3 E\epsilon^3 \\
&\quad -2\left\{2m_2 m_1 n^{-\alpha}\omega_2^2 + m_1 n^{-3\alpha/2}\omega_2^3 E\epsilon^3\right\} \\
&\quad -2\left\{2m_2 m_3 n^{-\alpha}\omega_2^2 + m_3 n^{-3\alpha/2}\omega_2^3 E\epsilon^3\right\} + 4\left\{n^{-\alpha} m_1 m_3 \omega_2^2\right\} \\
&= n^{-2\alpha}\omega_2^4\left(E\epsilon^4 - 1\right) + 2n^{-3\alpha/2}\omega_2^3 E\epsilon^3\left\{2m_2 - m_1 - m_3\right\} \\
&\quad +4n^{-\alpha}\omega_2^2\left(m_2 - m_3\right)\left(m_2 - m_1\right) \\
&= n^{-2\alpha}\omega_{t_{iK+1}}^4\left(E\epsilon^4 - 1\right) + O\left(n^{-3\alpha/2}\overline{n}^{-1}\right) + O\left(n^{-\alpha}\overline{n}^{-2}\right),
\end{aligned}$$

where we have used $\mathrm{var}\left(\varepsilon_i^2\right) = 4m_i^2 n^{-\alpha}\omega_i^2 + n^{-2\alpha}\omega_i^4\left(E\epsilon^4 - 1\right) + 4m_i n^{-3\alpha/2}\omega_i^3 E\epsilon^3$, $\mathrm{cov}\left(\varepsilon_{t_i}^2, \varepsilon_{t_{i+1}}\varepsilon_{t_i}\right) = 2m_{t_i}m_{t_{i+1}}n^{-\alpha}\omega_{t_i}^2 + m_{t_{i+1}}n^{-3\alpha/2}\omega_{t_i}^3 E\epsilon^3$ and $\mathrm{cov}\left(\varepsilon_1\varepsilon_2, \varepsilon_2\varepsilon_3\right) = n^{-\alpha}m_1 m_3\omega_2^2$.

The third component of $\mathrm{var}[u,u]^{n_1}$ is

$$\begin{aligned}
& 4\mathrm{var}[v,e]^{n_1} \\
&= 4\mathrm{var}\sum_{i=1}^{n_1-1}\left(v_{t_{iK+1}} - v_{t_{(i-1)K+1}}\right)\left(\varepsilon_{t_{iK+1}} - \varepsilon_{t_{(i-1)K+1}}\right) \\
&= 4\delta^2\gamma_n^2\mathrm{var}\sum_{i=1}^{n_1-1} Z_i\left(\varepsilon_{t_{iK+1}} - \varepsilon_{t_{(i-1)K+1}}\right),
\end{aligned}$$

where Z_i are 1-dependent $N\left(0, \frac{2T}{n}\right)$ r.v.'s with autocovariance $-\frac{1}{n}$, $Z \perp \varepsilon$

$$\begin{aligned}
&= 4\delta^2\gamma_n^2\sum_{i=1}^{n_1-1}\mathrm{var}\left[Z_i\Delta_K\varepsilon_{t_{iK+1}}\right] + 8\delta^2\gamma_n^2\sum_{i=2}^{n_1-1}\mathrm{cov}\left\{Z_i\Delta_K\varepsilon_{t_{iK+1}}, Z_{i-1}\Delta_K\varepsilon_{t_{(i-1)K+1}}\right\} \\
&= 4\delta^2\gamma_n^2\sum_{i=1}^{n_1-1} EZ_i^2 E\left(\varepsilon_{t_{iK+1}} - \varepsilon_{t_{(i-1)K+1}}\right)^2 + 8\delta^2\gamma_n^2\sum_{i=2}^{n_1-1}\left\{O\left(\frac{1}{n}\frac{K^2}{n^2}\right) + Tn^{-1-\alpha}\omega_{t_{(i-1)K+1}}^2\right\} \\
&= 24\delta^2 n^{-2\alpha} n_1 \textstyle\int \omega^2(u)\,du + O\left(n^{-2\alpha}\right) + O\left(n^{-\alpha}\frac{K}{n}\right).
\end{aligned}$$

PROOF OF LEMMA A5. We have

$$\begin{aligned}
E[X,u]^{avg} &= \frac{1}{K}\sum_{j=1}^{K} E[X,u]^{n_j} = \frac{1}{K}\sum_{j=1}^{K}\delta\gamma_n \sum_{i=1}^{n_j-1}\int\limits_{t_{iK+j-1}}^{t_{iK+j}} \sigma_t dt \\
&= \frac{\delta\gamma_n}{K}\sum_{i=1}^{n_j-1}\sum_{j=1}^{K}\int\limits_{t_{iK+j-1}}^{t_{iK+j}} \sigma_t dt = \frac{\delta\gamma_n}{K}\sum_{i=1}^{n_j-1}\int\limits_{t_{iK}}^{t_{(i+1)K}} \sigma_t dt \\
&= \frac{\delta\gamma_n}{K}\int\limits_{t_K}^{t_n} \sigma_t dt = \frac{\delta\gamma_n}{K}\int\limits_{0}^{1} \sigma_t dt + O\left(\frac{\gamma_n}{K}\frac{K}{n}\right) \\
&= \frac{\delta\gamma_n}{K}\int\limits_{0}^{1} \sigma_t dt + O\left(n^{-\frac{1+\alpha}{2}}\right).
\end{aligned}$$

As for the second part of Lemma A5, we first calculate $E[X,u]^{\{n\}}$,

$$\begin{aligned}
E[X,u]^{\{n\}} &= \sum_{i=K}^{n-K} E\left[\Delta X_{t_i}\Delta u_{t_i}\right] + \frac{1}{2}\left(\sum_{i=n-K+1}^{n} E\left[\Delta X_{t_i}\Delta u_{t_i}\right] + \sum_{i=1}^{K-1} E\left[\Delta X_{t_i}\Delta u_{t_i}\right]\right) \\
&= c\gamma_n \sum_{i=K}^{n-K}\int\limits_{t_{i-1}}^{t_i} \sigma_t dt + \frac{1}{2}c\gamma_n \sum_{i=n-K+1}^{n}\int\limits_{t_{i-1}}^{t_i} \sigma_t dt + \frac{1}{2}c\gamma_n \sum_{i=1}^{K-1}\int\limits_{t_{i-1}}^{t_i} \sigma_t dt \\
&= c\gamma_n \int\limits_{t_{K-1}}^{t_{n-K}} \sigma_t dt + \frac{1}{2}c\gamma_n\left(\int\limits_{t_{n-K}}^{t_n} \sigma_t dt + \int\limits_{t_0}^{t_{K-1}} \sigma_t dt\right) \\
&= c\gamma_n \int\limits_{0}^{1} \sigma_t dt - \frac{1}{2}c\gamma_n O\left(\overline{n}^{-1}\right).
\end{aligned}$$

Then,

$$\begin{aligned}
&\overline{n}^{1/2}E[X,u]^{avg} - \frac{\overline{n}^{3/2}}{n}E[X,u]^{\{n\}} \\
= {} & \overline{n}^{1/2}\frac{\delta\gamma_n}{K}\int\limits_{0}^{1} \sigma_t dt + \overline{n}^{1/2}O\left(n^{-\frac{1+\alpha}{2}}\right) - \frac{\overline{n}^{3/2}}{n}\delta\gamma_n\int\limits_{0}^{1} \sigma_t dt + \frac{\overline{n}^{3/2}}{n}O\left(\gamma_n\overline{n}^{-1}\right) \\
= {} & O\left(n^{-\frac{\alpha+\beta}{2}}\right) = o(1).
\end{aligned}$$

Proof of Lemma A6. We have

$$
\begin{aligned}
& \operatorname{var}[X,u]^{avg} \\
= & \operatorname{var}\left\{\frac{1}{K}\sum_{j=1}^{K}[X,u]^{n_j}\right\} \\
= & \frac{1}{K^2}\sum_{j=1}^{K}\operatorname{var}[X,u]^{n_j} + \frac{1}{K^2}\sum_{j\neq m}^{K}\sum_{m=1}^{K}\operatorname{cov}\left\{[X,u]^{n_j},[X,u]^{n_m}\right\} \\
= & \frac{1}{K^2}\sum_{j=1}^{K}\operatorname{var}[X,u]^{n_j} + O\left(n^{-1-\alpha}\overline{n}\right) + O\left(\overline{n}^{-1}n^{-1}\right) \qquad \text{(A.10)} \\
= & \frac{1}{K}\left(O\left(n^{-\alpha}\right) + O\left(\frac{1}{\overline{n}^2}\right)\right) + O\left(n^{-1-\alpha}\overline{n}\right) + O\left(\overline{n}^{-1}n^{-1}\right) \\
\sim & n^{-\beta-\alpha} + n^{-\beta-2+2\beta} + n^{-1-\alpha}n^{1-\beta} + n^{-1+\beta}n^{-1} \sim n^{\beta-2}.
\end{aligned}
$$

The above (A.10) follows by noticing that all covariance terms are of the same order, so we can explore the magnitude of one of them. Since we are looking at the magnitudes only, assume without loss of generality that $\delta = 1$. Then,

$$
\begin{aligned}
& \operatorname{cov}\left\{[X,u]^{\overline{n}}_{(1)},[X,u]^{\overline{n}}_{(2)}\right\} \\
= & \operatorname{cov}\left[\sum_{i=1}^{\overline{n}-1}\left(u_{t_{iK+1}} - u_{t_{(i-1)K+1}}\right)\left(X_{t_{iK+1}} - X_{t_{(i-1)K+1}}\right),\right. \\
& \left.\sum_{i=1}^{\overline{n}-1}\left(u_{t_{iK+2}} - u_{t_{(i-1)K+2}}\right)\left(X_{t_{iK+2}} - X_{t_{(i-1)K+2}}\right)\right]. \qquad \text{(A.11)}
\end{aligned}
$$

The latter can be easily shown to be $O\left(n^{-1-\alpha}\overline{n}\right) + O\left(\overline{n}^{-1}n^{-1}\right)$.

PROOF OF LEMMA A7. We have

$$
\begin{aligned}
& \overline{n}^{1/2}E[u,u]^{avg} - \frac{\overline{n}^{3/2}}{n}E\left[u,u\right]^{\{n\}} \\
= \; & \overline{n}^{1/2}\frac{1}{K}\sum_{j=1}^{K}\sum_{i=1}^{n_j-1}E\left[\left(u_{t_{iK+j}} - u_{t_{(i-1)K+j}}\right)^2\right] - \frac{\overline{n}^{3/2}}{n}E\left[u,u\right]^{\{n\}} \\
= \; & \frac{\overline{n}^{1/2}}{K}\sum_{i=1}^{n-K}E\left[\left(\varepsilon_{t_{i+K}} - \varepsilon_{t_i}\right)^2\right] && \text{(A.12)} \\
& -\frac{\overline{n}^{3/2}}{n}\frac{1}{2}\left\{\sum_{i=1}^{n-K}E\left[\left(\varepsilon_{t_{i+1}} - \varepsilon_{t_i}\right)^2\right] + \sum_{i=K}^{n-1}E\left[\left(\varepsilon_{t_{i+1}} - \varepsilon_{t_i}\right)^2\right]\right\} && \text{(A.13)} \\
= \; & n^{-\alpha}\frac{\overline{n}^{1/2}}{K}\sum_{i=1}^{n-K}\left(\omega_{t_{i+K}}^2 + \omega_{t_i}^2\right) && \text{(A.14)} \\
& -n^{-\alpha}\frac{\overline{n}^{3/2}}{2n}\left\{\sum_{i=1}^{n-K}\left(\omega_{t_{i+1}}^2 + \omega_{t_i}^2\right) + \sum_{i=K}^{n-1}\left(\omega_{t_{i+1}}^2 + \omega_{t_i}^2\right)\right\} && \text{(A.15)} \\
= \; & n^{-\alpha}\frac{\overline{n}^{1/2}}{K}\left\{\sum_{i=K+1}^{n}\omega_{t_i}^2 + \sum_{i=1}^{n-K}\omega_{t_i}^2 - \frac{1}{2}\sum_{i=2}^{n-K+1}\omega_{t_i}^2\right. \\
& \left. -\frac{1}{2}\sum_{i=1}^{n-K}\omega_{t_i}^2 - \frac{1}{2}\sum_{i=K+1}^{n}\omega_{t_i}^2 - \frac{1}{2}\sum_{i=K}^{n-1}\omega_{t_i}^2\right\} && \text{(A.16)} \\
\sim \; & n^{-\alpha}\frac{\overline{n}^{1/2}}{K} = o(1),
\end{aligned}
$$

where (A.12) follows because contributions from v are zero,

$$
\begin{aligned}
& \frac{\overline{n}^{1/2}}{K}(n-K)\,\delta^2\gamma_n^2\frac{1}{n} - \frac{1}{2}\left\{\frac{\overline{n}^{3/2}}{n}(n-K)\,\delta^2\gamma_n^2\frac{1}{n} + \frac{\overline{n}^{3/2}}{n}(n-K)\,\delta^2\gamma_n^2\frac{1}{n}\right\} \\
= \; & \frac{\overline{n}^{1/2}}{K}(n-K)\,\delta^2\gamma_n^2\frac{1}{n} - \frac{\overline{n}^{3/2}}{n}(n-K)\,\delta^2\gamma_n^2\frac{1}{n} = 0,
\end{aligned}
$$

and (A.14) follows because contributions from $m(.)$ are negligible,

$$
\begin{aligned}
& \frac{\overline{n}^{1/2}}{K}\sum_{i=1}^{n-K}O\left(\frac{1}{\overline{n}^2}\right) - \frac{\overline{n}^{3/2}}{n}\frac{1}{2}\left\{\sum_{i=1}^{n-K}O\left(\frac{1}{n^2}\right) + \sum_{i=K+1}^{n}O\left(\frac{1}{n^2}\right)\right\} \\
\sim \; & \frac{\overline{n}^{1/2}}{K}n\frac{1}{\overline{n}^2} + \frac{\overline{n}^{3/2}}{n}n\frac{1}{n^2} = \overline{n}^{-1/2} + \frac{\overline{n}^{3/2}}{n^2} = o(1).
\end{aligned}
$$

PROOF OF LEMMA A8. Using Lemma 4,

$$
\begin{aligned}
& \operatorname{var}[u,u]^{avg} \\
= \; & \operatorname{var}\left\{\frac{1}{K}\sum_{j=1}^{K}[u,u]^{n_j}\right\} = \frac{1}{K^2}\sum_{j=1}^{K}\operatorname{var}[u,u]^{n_j} \qquad \text{(A.17)} \\
= \; & \frac{1}{K}\left[\frac{n_1}{n^{2\alpha}}\left\{12\delta^4 + 4E\epsilon^4\int\omega^4(u)\,du + 24\delta^2\int\omega^2(u)\,du + O\left(\frac{n^{\alpha/2}}{n_1}\right) + O\left(\frac{n^{\alpha}}{n_1^2}\right)\right\}\right] \\
= \; & \frac{n_1}{Kn^{2\alpha}}\left\{12\delta^4 + 4E\epsilon^4\int\omega^4(u)\,du + 24\delta^2\int\omega^2(u)\,du + O\left(\frac{n^{\alpha/2}}{n_1}\right) + O\left(\frac{n^{\alpha}}{n_1^2}\right)\right\} \\
= \; & c^{-3}\overline{n}^{-1}\left\{12\delta^4 + 4E\epsilon^4\int\omega^4(u)\,du + 24\delta^2\int\omega^2(u)\,du + o(1)\right\} \text{ if } \beta = \frac{2}{3}(1-\alpha).
\end{aligned}
$$

In above, (A.17) follows because all covariance terms are zero. For example,

$$
\operatorname{cov}\left\{[u,u]^{n_1},[u,u]^{n_2}\right\} = \sum_{i=1}^{n_1-1}\sum_{j=1}^{n_2-1}\operatorname{cov}\left\{\left(u_{t_{iK+1}} - u_{t_{(i-1)K+1}}\right)^2, \left(u_{t_{jK+2}} - u_{t_{(j-1)K+2}}\right)^2\right\}.
$$

To show all terms in the summation above are zero, we do the calculation for the term with indices $(i=1, j=1)$

$$
\begin{aligned}
& \operatorname{cov}\left\{\left(u_{t_{K+1}} - u_{t_1}\right)^2, \left(u_{t_{K+2}} - u_{t_2}\right)^2\right\} \\
= \; & \operatorname{cov}\Big(\left[\delta\gamma_n\left(W_{t_{K+1}} - W_{t_K}\right) - \delta\gamma_n\left(W_{t_1} - W_{t_0}\right) + \left(\varepsilon_{t_{K+1}} - \varepsilon_{t_1}\right)\right]^2, \\
& \left[\delta\gamma_n\left(W_{t_{K+2}} - W_{t_{K+1}}\right) - \delta\gamma_n\left(W_{t_2} - W_{t_1}\right) + \left(\varepsilon_{t_{K+2}} - \varepsilon_{t_2}\right)\right]^2\Big) = 0
\end{aligned}
$$

as well as for the term with indices $(i=2, j=1)$

$$
\begin{aligned}
& \operatorname{cov}\left\{\left(u_{t_{2K+1}} - u_{t_{K+1}}\right)^2, \left(u_{t_{K+2}} - u_{t_2}\right)^2\right\} \\
= \; & \operatorname{cov}\Big(\left[\delta\gamma_n\left(W_{t_{2K+1}} - W_{t_{2K}}\right) - \delta\gamma_n\left(W_{t_{K+1}} - W_{t_K}\right) + \left(\varepsilon_{t_{2K+1}} - \varepsilon_{t_{K+1}}\right)\right]^2, \\
& \left[\delta\gamma_n\left(W_{t_{K+2}} - W_{t_{K+1}}\right) - \delta\gamma_n\left(W_{t_2} - W_{t_1}\right) + \left(\varepsilon_{t_{K+2}} - \varepsilon_{t_2}\right)\right]^2\Big) = 0.
\end{aligned}
$$

PROOF LEMMA B1. We have

$$
\begin{aligned}
&\operatorname{cov}\left([X,X]^{\overline{n}}_{avg},[u,u]^{\overline{n}}_{avg}\right)\\
= &\operatorname{cov}\left(\frac{1}{K}\sum_{j=1}^{K}[X,X]^{n_j},\frac{1}{K}\sum_{j=1}^{K}[u,u]^{n_j}\right)\\
= &\operatorname{cov}\left(\frac{1}{K}\sum_{j=1}^{K}\sum_{i=1}^{n_j-1}\left(X_{t_{iK+j}}-X_{t_{(i-1)K+j}}\right)^2,\frac{1}{K}\sum_{j=1}^{K}\sum_{i=1}^{n_j-1}\left(u_{t_{iK+j}}-u_{t_{(i-1)K+j}}\right)^2\right)\\
= &\operatorname{cov}\left(\frac{1}{K}\sum_{j=1}^{K}\sum_{i=1}^{n_j-1}a_{ij},\frac{1}{K}\sum_{j=1}^{K}\sum_{i=1}^{n_j-1}\left(\gamma_n b_{1,ij}-\gamma_n b_{2,ij}+b_{3,ij}\right)^2\right)\\
= &\operatorname{cov}\left(\frac{1}{K}\sum_{j=1}^{K}\sum_{i=1}^{n_j-1}a_{ij},\right.\\
&\left.\frac{1}{K}\sum_{j=1}^{K}\sum_{i=1}^{n_j-1}\left(\gamma_n^2 b_{1,ij}^2+\gamma_n^2 b_{2,ij}^2+b_{3,ij}^2-2\gamma_n^2 b_{1,ij}b_{2,ij}+2\gamma_n b_{1,ij}b_{3,ij}-2\gamma_n b_{2,ij}b_{3,ij}\right)\right)\\
= &c_1+c_2+c_3+c_4+c_5+c_6\\
= &O\left(K^{-1}n^{-\alpha}\right)=o\left(\overline{n}^{-1}\right)\text{ as }-\beta-\alpha<-(1-\beta)\text{ holds if }\beta=\frac{2}{3}(1-\alpha)
\end{aligned}
$$

where $b_{1,ij}=W_{t_{iK+j}}-W_{t_{iK+j-1}}$, $b_{2,ij}=W_{t_{(i-1)K+j}}-W_{t_{(i-1)K+j-1}}$, $b_{3,ij}=\varepsilon_{t_{iK+j}}-\varepsilon_{t_{(i-1)K+j}}$.

The last line follows because $c1\sim c2\sim K^{-2}$ and $c_3=c_4=c_5=c_6=0$ by properties of normal random variables.

PROOF OF LEMMA B2. First, note that $\operatorname{cov}\left([X,X]^{\overline{n}}_{avg},[u,u]^{n}\right)$ is of the same order as

$\operatorname{cov}\left([X,X]^{\overline{n}}_{avg},[u,u]^{\{n\}}\right)$ since $[u,u]^{n}-[u,u]^{\{n\}}$ is of smaller order than $[u,u]^{\{n\}}$. Also, notice that $\operatorname{cov}\left([X,X]^{\overline{n}}_{avg},[u,u]^{n}\right)$ has to be of the same order as $\operatorname{cov}\left([X,X]^{\overline{n}}_{avg},[u,u]^{\overline{n}}_{avg}\right)$ by similarity in construction of $[u,u]^{n}$ and $[u,u]^{\overline{n}}_{avg}$. Hence, $\operatorname{cov}\left([X,X]^{\overline{n}}_{avg},[u,u]^{\{n\}}\right)=o\left(\overline{n}^{-1}\right)$ by Lemma B1.

PROOF OF LEMMA B3. We need to prove here that

$$\operatorname{cov}\left([u,u]^{\{n\}},[u,u]^{avg}\right)=\frac{n}{\bar{n}^2}c^{-3}\left\{8\delta^4+4\left(E\epsilon^4-1\right)\sigma_\epsilon^4+16\delta^2\int\omega^2(u)\,du+o(1)\right\}.$$

We will calculate the expression for $\operatorname{cov}\left([u,u]^{avg},[u,u]^n\right)$. It has the same leading term as

$\operatorname{cov}\left([u,u]^{avg},[u,u]^{\{n\}}\right)$ since $[u,u]^n-[u,u]^{\{n\}}$ is of smaller order than $[u,u]^{\{n\}}$.

$$\begin{aligned}\operatorname{cov}\left([u,u]^{avg},[u,u]^n\right) &= \operatorname{cov}\left(\frac{1}{K}\sum_{j=1}^{K}\sum_{i=1}^{n_j-1}\left(u_{t_{iK+j}}-u_{t_{(i-1)K+j}}\right)^2,\sum_{i=1}^{n-1}\left(u_{t_{i+1}}-u_{t_i}\right)^2\right)\\ &= \operatorname{cov}\left(\frac{1}{K}\sum_{i=1}^{n-K}\left(u_{t_{i+K}}-u_{t_i}\right)^2,\sum_{i=1}^{n-1}\left(u_{t_{i+1}}-u_{t_i}\right)^2\right)\end{aligned}$$

$$=\operatorname{cov}\left(\frac{1}{K}\sum_{i=1}^{n-K}\left(\underbrace{v_{t_{i+K}}}_{a_1}\underbrace{-v_{t_i}}_{a_2}+\underbrace{\varepsilon_{t_{i+K}}-\varepsilon_{t_i}}_{a_3}\right)^2,\sum_{i=1}^{n-1}\left(\underbrace{v_{t_{i+1}}}_{b_1}\underbrace{-v_{t_i}}_{b_2}+\underbrace{\varepsilon_{t_{i+1}}-\varepsilon_{t_i}}_{b_3}\right)^2\right)$$

$$\begin{aligned}&= \frac{1}{K}\operatorname{cov}\left(a_1^2+a_2^2+a_3^2+2a_1a_2+2a_1a_3+2a_2a_3,\right.\\ &\qquad \left.b_1^2+b_2^2+b_3^2+2b_1b_2+2b_1b_3+2b_2b_3\right)\\ &= \frac{1}{K}\operatorname{cov}\left(a_1^2+a_2^2+a_3^2,b_1^2+b_2^2+b_3^2\right)\\ &\quad+\frac{2}{K}\operatorname{cov}\left(a_1^2+a_2^2+a_3^2,b_1b_2+b_1b_3+b_2b_3\right)\\ &\quad+\frac{2}{K}\operatorname{cov}\left(a_1a_2+a_1a_3+a_2a_3,b_1^2+b_2^2+b_3^2\right)\\ &\quad+\frac{4}{K}\operatorname{cov}\left(a_1a_2+a_1a_3+a_2a_3,b_1b_2+b_1b_3+b_2b_3\right)\end{aligned}$$

Denote the terms in last four lines by

$$
\begin{aligned}
& \operatorname{cov}\left([u,u]^{avg},[u,u]^{n}\right) \\
= \; & B3_1 + B3_2 + B3_3 + B3_4 \\
= \; & B3_1 + 0 + 0 + \left\{16\delta^2 c^{-1} n^{1-2\alpha-\beta} \int \omega^2(u)\,du + o\left(n^{1-2\alpha-\beta}\right)\right\} \\
= \; & 8\delta^4 K^{-1} n^{1-2\alpha} + 4\left(E\epsilon^4 - 1\right) K^{-1} n^{1-2\alpha} \int_0^1 \omega_u^4 du + o\left(K^{-1} n^{1-2\alpha}\right) + O\left(n^{-\alpha}\right) \\
& + K^{-1} n^{1-2\alpha} \left\{16\delta^2 \int \omega^2(u)\,du + o\left(n^{1-2\alpha-\beta}\right)\right\} \\
= \; & K^{-1} n^{1-2\alpha} \left\{8\delta^4 + 4\left(E\epsilon^4 - 1\right)\sigma_\epsilon^4 + 16\delta^2 \int \omega^2(u)\,du + o(1) + O\left(n^{\alpha+\beta-1}\right)\right\} \\
= \; & \frac{n}{\overline{n}^2} c^{-3} \left\{8\delta^4 + 4\left(E\epsilon^4 - 1\right)\sigma_\epsilon^4 + 16\delta^2 \int \omega^2(u)\,du + o(1)\right\} \text{ if } \beta = \frac{2}{3}(1-\alpha).
\end{aligned}
$$

This result is the same as in Zhang et al. (2005) paper, where the covariance is, apart from normalisation factor, $\operatorname{cov}\left([\epsilon,\epsilon]^{avg},[\epsilon,\epsilon]^{n}\right) = 4\operatorname{var}\left(\epsilon^2\right) = 4\left(E\epsilon^4 - 1\right).$

To obtain the expression for the $B3_1$ term, note that the terms $\operatorname{cov}\left(a_1^2, b_1^2\right)$, $\operatorname{cov}\left(a_1^2, b_2^2\right)$, $\operatorname{cov}\left(a_2^2, b_1^2\right)$, and $\operatorname{cov}\left(a_2^2, b_2^2\right)$ are all equal to $2\delta^4\gamma_n^4 n^{-1} + o\left(n^{1-2\alpha}\right)$, and also $\operatorname{cov}\left(a_1^2, b_3^2\right) = \operatorname{cov}\left(a_2^2, b_3^2\right) = \operatorname{cov}\left(a_3^2, b_1^2\right) = \operatorname{cov}\left(a_3^2, b_2^2\right) = 0$. The final term in $B3_1$ is

$$
\begin{aligned}
\operatorname{cov}\left(a_3^2, b_3^2\right) &= \operatorname{cov}\left(\sum_{i=1}^{n-K}\left(\varepsilon_{t_{i+K}} - \varepsilon_{t_i}\right)^2, \sum_{i=1}^{n-1}\left(\varepsilon_{t_{i+1}} - \varepsilon_{t_i}\right)^2\right) \\
&= 4\left(E\epsilon^4 - 1\right) n^{1-2\alpha} \int_0^1 \omega_u^4 du + o\left(n^{1-2\alpha}\right) + O\left(Kn^{-\alpha}\right),
\end{aligned}
$$

using similar steps as in Lemma A4. Hence,

$$
\begin{aligned}
B3_1 &= \frac{1}{K}\operatorname{cov}\left(a_1^2 + a_2^2 + a_3^2, b_1^2 + b_2^2 + b_3^2\right) \\
&= 8\delta^4 K^{-1} n^{1-2\alpha} + 4\left(E\epsilon^4 - 1\right) K^{-1} n^{1-2\alpha} \int_0^1 \omega_u^4 du + o\left(K^{-1} n^{1-2\alpha}\right) + O\left(n^{-\alpha}\right).
\end{aligned}
$$

The next two terms are zero, $B3_2 = B3_3 = 0$.

Finally, we show how to obtain $B3_4 = 16\delta^2 c^{-1} n^{1-2\alpha-\beta} \int \omega^2(u)\,du + o\left(n^{1-2\alpha-\beta}\right)$.

We have

$$
\begin{aligned}
& B3_4 \\
= \; & \frac{4}{K}\mathrm{cov}\left(a_1a_2 + a_1a_3 + a_2a_3, b_1b_2 + b_1b_3 + b_2b_3\right) \\
= \; & 0+0+0+ \\
& 0+\frac{4}{K}\left\{O\left(\frac{\gamma_n^2}{\overline{n}^2}\right)+\delta^2 n^{-\alpha}\gamma_n^2\frac{1}{n}\sum_{i=K}^{n-1}\omega_{t_{i+1}}^2\right\}+\frac{4}{K}\left\{O\left(\frac{\gamma_n^2}{\overline{n}^2}\right)+\delta^2 n^{-\alpha}\gamma_n^2\frac{1}{n}\sum_{i=K}^{n-2}\omega_{t_{i+1}}^2\right\}+ \\
& 0+\frac{4}{K}\left\{O\left(\frac{\gamma_n^2}{\overline{n}^2}\right)+\delta^2 n^{-\alpha}\gamma_n^2\frac{1}{n}\sum_{i=1}^{n-K-1}\omega_{t_{i+1}}^2\right\}+\frac{4}{K}\left\{O\left(\frac{\gamma_n^2}{\overline{n}^2}\right)+\delta^2 n^{-\alpha}\gamma_n^2\frac{1}{n}\sum_{i=0}^{n-K-1}\omega_{t_{i+1}}^2\right\}
\end{aligned}
$$

$$
\begin{aligned}
= \; & 16\delta^2 c^{-1} n^{1-2\alpha-\beta}\int\omega^2(u)\,du + o\left(n^{1-2\alpha-\beta}\right)+O\left(\frac{\gamma_n^2}{K\overline{n}^2}\right) \\
= \; & 16\delta^2 c^{-1} n^{1-2\alpha-\beta}\int\omega^2(u)\,du + o\left(n^{1-2\alpha-\beta}\right)+O\left(n^{1-\alpha}n^{-\beta}n^{-2(1-\beta)}\right) \\
= \; & 16\delta^2 c^{-1} n^{1-2\alpha-\beta}\int\omega^2(u)\,du + o\left(n^{1-2\alpha-\beta}\right).
\end{aligned}
$$

The first equality above follows because $\mathrm{cov}\left(a_1a_2, b_1b_2\right) = \mathrm{cov}\left(a_1a_2, b_1b_3\right) = \mathrm{cov}\left(a_1a_2, b_2b_3\right) = \mathrm{cov}\left(a_1a_3, b_1b_2\right) = \mathrm{cov}\left(a_2a_3, b_1b_2\right) = 0$ and we have:

$$
\begin{aligned}
& \mathrm{cov}\left(a_1a_3, b_1b_3\right) \\
= \; & \delta^2\gamma_n^2\mathrm{cov}\left(\sum_{i=1}^{n-K}\left(W_{t_{i+K}}-W_{t_{i+K-1}}\right)\left(\varepsilon_{t_{i+K}}-\varepsilon_{t_i}\right),\sum_{i=1}^{n-1}\left(W_{t_{i+1}}-W_{t_i}\right)\left(\varepsilon_{t_{i+1}}-\varepsilon_{t_i}\right)\right) \\
= \; & O\left(\frac{\gamma_n^2}{\overline{n}^2}\right)+\delta^2 n^{-\alpha}\gamma_n^2\frac{1}{n}\sum_{i=K}^{n-1}\omega_{t_{i+1}}^2,
\end{aligned}
$$

$$
\begin{aligned}
& \operatorname{cov}\left(a_1 a_3, b_2 b_3\right) \\
= & \ \delta^2 \gamma_n^2 \operatorname{cov}\left(\sum_{i=1}^{n-K}\left(W_{t_{i+K}}-W_{t_{i+K-1}}\right)\left(\varepsilon_{t_{i+K}}-\varepsilon_{t_i}\right), \sum_{i=1}^{n-1}-\left(W_{t_i}-W_{t_{i-1}}\right)\left(\varepsilon_{t_{i+1}}-\varepsilon_{t_i}\right)\right) \\
= & \ O\left(\frac{\gamma_n^2}{\bar{n}^2}\right)+\delta^2 n^{-\alpha} \gamma_n^2 \frac{1}{n} \sum_{i=K}^{n-2} \omega_{t_{i+1}}^2,
\end{aligned}
$$

$$
\begin{aligned}
& \operatorname{cov}\left(a_2 a_3, b_1 b_3\right) \\
= & \ \delta^2 \gamma_n^2 \operatorname{cov}\left(\sum_{i=1}^{n-K}-\left(W_{t_i}-W_{t_{i-1}}\right)\left(\varepsilon_{t_{i+K}}-\varepsilon_{t_i}\right), \sum_{i=1}^{n-1}\left(W_{t_{i+1}}-W_{t_i}\right)\left(\varepsilon_{t_{i+1}}-\varepsilon_{t_i}\right)\right) \\
= & \ O\left(\frac{\gamma_n^2}{\bar{n}^2}\right)+\delta^2 n^{-\alpha} \gamma_n^2 \frac{1}{n} \sum_{i=1}^{n-K-1} \omega_{t_{i+1}}^2,
\end{aligned}
$$

$$
\begin{aligned}
& \operatorname{cov}\left(a_2 a_3, b_2 b_3\right) \\
= & \ \delta^2 \gamma_n^2 \operatorname{cov}\left(\sum_{i=1}^{n-K}-\left(W_{t_i}-W_{t_{i-1}}\right)\left(\varepsilon_{t_{i+K}}-\varepsilon_{t_i}\right), \sum_{i=1}^{n-1}-\left(W_{t_i}-W_{t_{i-1}}\right)\left(\varepsilon_{t_{i+1}}-\varepsilon_{t_i}\right)\right) \\
= & \ O\left(\frac{\gamma_n^2}{\bar{n}^2}\right)+\delta^2 n^{-\alpha} \gamma_n^2 \frac{1}{n} \sum_{i=0}^{n-K-1} \omega_{t_{i+1}}^2 .
\end{aligned}
$$

A.3 Tables and Figures

$n \searrow \alpha$	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
195	96	186	145	120	145	114	95	78	65	54	N/A
390	94	135	110	200	156	128	143	111	89	71	59
780	67	90	108	137	107	181	151	162	119	100	76
1560	55	74	67	86	94	125	205	161	119	125	92
4680	48	47	56	58	74	96	99	117	201	144	151
5850	44	51	57	57	66	81	76	135	98	160	163
7800	45	46	52	53	68	70	90	94	109	175	134
11700	40	44	45	52	53	59	81	78	141	208	148
23400	36	40	43	46	49	58	61	79	106	123	196

Table A.1: IQR percentage error with $K = (2V_2/V_1)^{1/3} n^{\frac{2}{3}(1-\alpha)}$

$n \searrow \alpha$	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
195	3	2	2	2	1	1	1	1	1	1	0
390	4	3	3	2	2	2	1	1	1	1	1
780	7	5	4	3	3	2	2	1	1	1	1
1560	11	8	7	5	4	3	2	2	2	1	1
4680	22	17	13	10	7	5	4	3	2	2	1
5850	26	19	14	11	8	6	5	3	3	2	1
7800	31	23	17	13	9	7	5	4	3	2	2
11700	41	30	22	16	12	9	6	5	3	2	2
23400	65	47	33	24	17	12	9	6	4	3	2

Table A.2: $K = (2V_2/V_1)^{1/3} n^{\frac{2}{3}(1-\alpha)}$

$n \diagdown \alpha$	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
195	-21	-16	-13	-7	-7	-3	-1	4	8	13	13
390	-15	-12	-7	-3	-3	1	3	6	7	12	14
780	-13	-11	-4	-2	0	0	4	5	6	11	14
1560	-9	-7	-2	-1	1	3	5	7	8	13	12
4680	-5	-3	-1	-2	1	0	3	5	6	7	11
5850	-4	-3	1	3	5	5	2	4	8	8	8
7800	-2	-2	0	1	3	2	5	3	6	8	10
11700	-3	0	0	2	2	5	4	2	6	3	8
23400	-2	1	2	1	3	4	2	6	6	6	8

Table A.3: IQR percentage error with $K = n^{\frac{2}{3}(1-\alpha)}$

$n \diagdown \alpha$	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
195	34	28	24	20	17	14	12	10	8	7	6
390	53	44	36	29	24	20	16	13	11	9	7
780	85	68	54	44	35	28	22	18	14	11	9
1560	135	105	82	64	50	39	31	24	19	15	12
4680	280	211	159	120	91	68	52	39	29	22	17
5850	325	243	182	136	102	76	57	43	32	24	18
7800	393	292	216	161	119	88	66	49	36	27	20
11700	515	377	276	202	148	108	79	58	42	31	23
23400	818	585	418	299	214	153	109	78	56	40	29

Table A.4: $K = n^{\frac{2}{3}(1-\alpha)}$

$n \setminus \alpha$	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	K
195	-23	-23	-24	-23	-23	-21	-23	-24	-23	-24	-23	34
390	-17	-19	-19	-17	-19	-20	-18	-16	-16	-18	-18	53
780	-14	-15	-12	-15	-14	-12	-15	-15	-16	-14	-13	85
1560	-12	-9	-10	-10	-12	-11	-11	-9	-11	-12	-9	135
4680	-7	-2	-7	-5	-5	-7	-6	-5	-5	-6	-5	280
5850	-6	-6	-6	-6	-6	-6	-5	-7	-6	-5	-4	325
7800	-5	-6	-4	-4	-3	-4	-5	-4	-5	-6	-5	393
11700	-2	-6	-3	-3	-3	-4	-2	-5	-6	-2	-3	515
23400	-2	-2	-3	-2	-1	-2	-1	-3	-4	-2	-4	818

Table A.5: IQR percentage error with $K = n^{\frac{2}{3}}$

$n \setminus \alpha$	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	K^{BR}
195	55	46	34	29	27	21	22	19	16	18	15	6
390	67	49	37	28	23	20	17	18	15	15	14	8
780	94	65	48	32	26	22	19	16	16	14	12	10
1560	124	81	54	36	27	24	15	14	14	13	13	13
4680	243	146	91	54	34	24	18	16	12	14	8	18
5850	263	155	92	53	35	24	18	11	11	11	12	20
7800	300	182	97	60	33	26	15	13	10	11	9	22
11700	381	223	125	68	39	24	17	11	12	9	8	25
23400	539	305	163	86	47	28	15	13	8	8	8	32

Table A.6: IQR percentage error with $K^{BR} = \phi = \left(\frac{3RV^2}{2RQ}\right)^{1/3} n^{1/3}$

δ^2/δ_b^2	$corr(\Delta X_{t_i}, \Delta u_{t_i})$	IQR error
0.01	-0.0010	0.0133
0.05	-0.0051	0.0128
0.1	-0.0102	0.0049
0.25	-0.0254	0.0182
0.5	-0.0506	0.0037
1	-0.1000	0.0136
2	-0.1909	0.0100
4	-0.3280	0.0090
10	-0.4869	0.0130
20	-0.5351	0.0105

Table A.7: Effect of δ^2 on the estimates $\widehat{QV}_x$

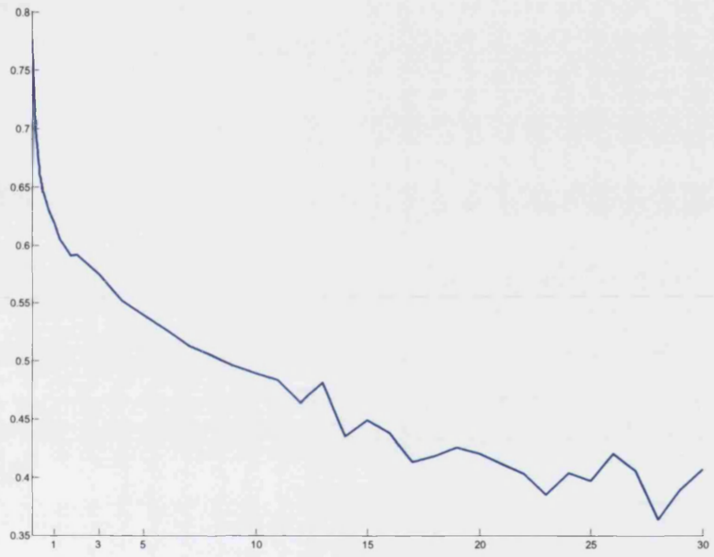


Figure A.1: *Volatility signature plot for IBM transactions data, year 2005. The scale on the X axis is the frequency of interpolated calendar time observations, in minutes. Y axis denotes average of daily RV using calendar time data at frequency specified by the x axis.*

Appendix B

Appendices for Chapter 2

B.1 Proofs of Chapter 2

Since $\{\sigma_t\}, \{\widetilde{\sigma}_t\}, \{\mu_t\}$ and $\{\widetilde{\mu}_t\}$ are locally bounded, it can be assumed, without loss of generality, that they are uniformly bounded by C_σ (see Barndorff-Nielsen, Graversen, Jacod, and Shephard (2006), Section 3). We use C to denote a generic constant that is different from line to line.

B.1.1 Proof of Proposition 1

By Cauchy-Schwarz and Burkholder-Davis-Gundy inequality (Revuz and Yor, 2005, p. 160),

$$
\begin{aligned}
\mathrm{E}\widehat{\theta}_{n,m,l} &= \sum_{i=m(l-1)}^{ml} \mathrm{E}\left(X_{i/n}-X_{(i-1)/n}\right)^2 \leq C \sum_{i=m(l-1)}^{ml} \int_{(i-1)/n}^{i/n} \sigma_u^4 du \leq CC_\sigma \frac{m}{n}, \\
\mathrm{Var}\widehat{\theta}_{n,m,l} &= \sum_{i'=m(l-1)}^{ml} \sum_{i=m(l-1)}^{ml} \mathrm{Cov}\left[\left(X_{i/n}-X_{(i-1)/n}\right)^2, \left(X_{i'/n}-X_{(i'-1)/n}\right)^2\right] \\
&\leq \sum_{i'=m(l-1)}^{ml} \sum_{i=m(l-1)}^{ml} \mathrm{E}\left[\left(X_{i/n}-X_{(i-1)/n}\right)^2 \left(X_{i'/n}-X_{(i'-1)/n}\right)^2\right] \\
&\leq \sum_{i'=m(l-1)}^{ml} \sum_{i=m(l-1)}^{ml} \mathrm{E}\left[\left(X_{i/n}-X_{(i-1)/n}\right)^4\right]^{1/2} \mathrm{E}\left[\left(X_{i'/n}-X_{(i'-1)/n}\right)^4\right]^{1/2} \\
&\leq C \sum_{i'=m(l-1)}^{ml} \sum_{i=m(l-1)}^{ml} \mathrm{E}\left[\left(\int_{(i-1)/n}^{i/n} \sigma_u^4 du\right)^2\right]^{1/2} \mathrm{E}\left[\left(\int_{(i'-1)/n}^{i'/n} \sigma_u^4 du\right)^2\right]^{1/2} \\
&\leq CC_\sigma m^2 n^{-2}
\end{aligned}
$$

for some constant C. Hence,

$$\widehat{\theta}_{n,m,l} = O_p\left(\frac{m}{n}\right)$$

and

$$
\begin{aligned}
\widehat{V}_{PR} &= m \times \frac{1}{K} \sum_{l=1}^{K} \left(\widehat{\theta}_{n,m,l} - \widehat{\theta}_n\right)^2 \\
&= m\widehat{\theta}_n - 2\widehat{\theta}_n m \times \frac{1}{K} \sum_{l=1}^{K} \widehat{\theta}_{n,m,l} + m \times \frac{1}{K} \sum_{l=1}^{K} \left(\widehat{\theta}_{n,m,l}\right)^2 \\
&= m\widehat{\theta}_n - 2\frac{m}{K}\widehat{\theta}_n^2 + \frac{m}{K} \sum_{l=1}^{K} \left(\widehat{\theta}_{n,m,l}\right)^2 \\
&= m\widehat{\theta}_n + o_p(m).
\end{aligned}
$$

The result now follows by consistency of $\widehat{\theta}_n$ for θ. ∎

B.1.2 Proof of Proposition 2

Before proceeding to the main proof, we state two useful inequalities that hold when X and its volatility are Brownian semimartingales. First, for any $q > 0$

$$\mathrm{E}\left(\left|\sigma_{t+s} - \sigma_t\right|^q \left|\mathcal{F}_t\right.\right) \leq C s^{q/2}. \tag{B.1}$$

This holds because

$$\begin{aligned}
\mathrm{E}\left(\left|\sigma_{t+s} - \sigma_t\right|^q \left|\mathcal{F}_t\right.\right) &= \mathrm{E}\left(\left|\int_t^{s+t} \widetilde{\mu}_u du + \int_t^{s+t} \widetilde{\sigma}_u d\widetilde{W}_u\right|^q \left|\mathcal{F}_t\right.\right) \\
&\leq \mathrm{E}\left(\left|\int_t^{s+t} \widetilde{\mu}_u du\right|^q \left|\mathcal{F}_t\right.\right) + \mathrm{E}\left(\left|\int_t^{s+t} \widetilde{\sigma}_u d\widetilde{W}_u\right|^q \left|\mathcal{F}_t\right.\right) \\
&\leq C s^q + C\mathrm{E}\left(\left|\int_t^{s+t} \widetilde{\sigma}_u^2 du\right|^{q/2} \left|\mathcal{F}_t\right.\right) \\
&\leq C s^{q/2}
\end{aligned}$$

were Davis-Burkholder-Gundy inequality (Revuz and Yor, 2005, p. 160) is used to obtain the second transition.

The second inequality is as follows, see Jacod (2007). For for all $q > 1$,

$$\mathrm{E}\left[\left|\mathcal{X}_{k.i}\right|^q \left|\mathcal{F}_{\frac{(k-1)m+i-1}{n}}\right.\right] \leq C\left(\frac{1}{n}\right)^{1\wedge q/2} \tag{B.2}$$

where

$$\begin{aligned}
\mathcal{X}_{k.i} &= \sqrt{n}\left[\sigma_{\frac{m(k-1)}{n}} \Delta W_{\frac{(k-1)m+i}{n}} - \Delta X_{\frac{(k-1)m+i}{n}}\right] \\
&= \sqrt{n} \int_{[(k-1)m+i-1]/n}^{[(k-1)m+i]/n} \left(\mu_u du + \left(\sigma_u - \sigma_{\frac{m(k-1)}{n}}\right) dW_u\right).
\end{aligned}$$

Introduce the following notation,

$$\widehat{V}_{sub}^{DISCR} = \frac{1}{K}\sum_{k=1}^{K} 2\sigma^4_{\frac{k-1}{K}} \qquad \mathcal{E}\left(\widehat{V}\right)^{DISCR} = \frac{J}{K}\sum_{k=1}^{K} \mathrm{E}\left[\gamma_k^{DISCR}\left|\mathcal{F}_{\frac{k-1}{K}}\right.\right]$$

$$\mathcal{E}\left(\widehat{V}\right) = \frac{J}{K}\sum_{k=1}^{K} \mathrm{E}\left[\gamma_k\left|\mathcal{F}_{\frac{k-1}{K}}\right.\right] \qquad \widehat{\alpha}_k^{short} = \frac{n}{J}\sum_{i=1}^{J}\sigma^2_{\frac{m(k-1)}{n}}\left(W_{\frac{(k-1)m+i}{n}} - W_{\frac{(k-1)m+i-1}{n}}\right)^2$$

$$\gamma_k = \left(\widehat{\theta}_k^{short} - \widehat{\theta}_k^{long}\right)^2 \qquad \widehat{\alpha}_k^{long} = \frac{n}{m}\sum_{i=1}^{m}\sigma^2_{\frac{m(k-1)}{n}}\left(W_{\frac{(k-1)m+i}{n}} - W_{\frac{(k-1)m+i-1}{n}}\right)^2$$

$$\gamma_k^{DISCR} = \left(\widehat{\alpha}_k^{short} - \widehat{\alpha}_k^{long}\right)^2$$

We want to show

$$\widehat{V}_{sub} = \frac{1}{K}\sum_{k=1}^{K}\left(\widehat{\theta}_k^{slow} - \widehat{\theta}_k^{fast}\right)^2 \xrightarrow{p} V = 2\int_0^1 \sigma_u^4 du.$$

First, by Riemann integrability of σ,

$$V^{DISCR} \xrightarrow{p} V = 2\int_0^1 \sigma_u^4 du.$$

To prove Proposition 2, proceed in three steps. Prove $\widehat{V} - \mathcal{E}\left(\widehat{V}\right) \xrightarrow{p} 0$, then $\mathcal{E}\left(\widehat{V}\right)^{DISCR} - \mathcal{E}\left(\widehat{V}\right) \xrightarrow{p} 0$, and finally $\mathcal{E}\left(\widehat{V}\right)^{DISCR} - V^{DISCR} \xrightarrow{p} 0$.

The first step is to show

$$\widehat{V} - \mathcal{E}\left(\widehat{V}\right) = \frac{J}{K}\sum_{k=1}^{K}\left(\gamma_k - \mathrm{E}\left[\gamma_k\left|\mathcal{F}_{\frac{k-1}{K}}\right.\right]\right) \xrightarrow{p} 0.$$

By Lenglart's inequality (see e.g. Podolskij 2006), it is sufficient to show that

$$\sum_{k=1}^{K}\mathrm{E}\left[\left|\frac{J}{K}\gamma_k\right|^2\left|\mathcal{F}_{\frac{k-1}{K}}\right.\right] \xrightarrow{p} 0.$$

We have

$$
\begin{aligned}
& \mathrm{E}\left[\left|\frac{J}{K}\gamma_k\right|^2 \middle| \mathcal{F}_{\frac{k-1}{K}}\right] \\
= & \frac{J^2}{K^2}\mathrm{E}\left[\left\{\frac{n}{J}\sum_{i=1}^{J}\left(X_{\frac{(k-1)m+i}{n}} - X_{\frac{(k-1)m+i-1}{n}}\right)^2 - \frac{n}{m}\sum_{i=1}^{m}\left(X_{\frac{(k-1)m+i}{n}} - X_{\frac{(k-1)m+i-1}{n}}\right)^2\right\}^4 \middle| \mathcal{F}_{\frac{k-1}{K}}\right] \\
\leq & C\frac{J^2}{K^2} = \frac{J^2m^2}{n^2}
\end{aligned}
$$

for some constant C not depending on k, by repeated use of Cauchy-Schwarz inequality and

$$
\mathrm{E}\left[\left|X_{\frac{(k-1)m+i}{n}} - X_{\frac{(k-1)m+i-1}{n}}\right|^q \middle| \mathcal{F}_{\frac{k-1}{K}}\right] \leq C_q\left(\frac{1}{n}\right)^{q/2}
$$

for all $q > 0$, $i = 1, ..., m$, and C_q some constant depending on q only. Hence,

$$
\sum_{k=1}^{K}\mathrm{E}\left[\left|\frac{J}{K}\gamma_k\right|^2 \middle| \mathcal{F}_{\frac{k-1}{K}}\right] \leq C\frac{J^2}{K} = \frac{mJ^2}{n}.
$$

The first step is thus proved, provided $mJ^2n^{-1} \to 0$.

Second step is to show

$$
\mathcal{E}\left(\widehat{V}\right)^{DISCR} - \mathcal{E}\left(\widehat{V}\right) = \frac{J}{K}\sum_{k=1}^{K}\mathrm{E}\left[\gamma_k^{DISCR} - \gamma_k \middle| \mathcal{F}_{\frac{k-1}{K}}\right] \xrightarrow{p} 0.
$$

We have

$$
\begin{aligned}
& \mathrm{E}\left[\left|\gamma_k^{DISCR} - \gamma_k\right| \middle| \mathcal{F}_{\frac{k-1}{K}}\right] \\
= & \mathrm{E}\left[\left|\widehat{\alpha}_k^{long} - \widehat{\alpha}_k^{short} + \widehat{\theta}_k^{long} - \widehat{\theta}_k^{short}\right| \left|\left\{\widehat{\alpha}_k^{long} - \widehat{\theta}_k^{long}\right\} - \left\{\widehat{\alpha}_k^{short} - \widehat{\theta}_k^{short}\right\}\right| \middle| \mathcal{F}_{\frac{k-1}{K}}\right] \\
= & \mathrm{E}_k\left[\left|\widehat{\alpha}_k^{long} - \widehat{\alpha}_k^{short} + \widehat{\theta}_k^{long} - \widehat{\theta}_k^{short}\right| \times \right. \\
& \left|\frac{n}{m}\sum_{i=1}^{m}\left[\sigma^2_{\frac{m(k-1)}{n}}\left(\Delta W_{\frac{(k-1)m+i}{n}}\right)^2 - \left(\Delta X_{\frac{(k-1)m+i}{n}}\right)^2\right]\right. \\
& \left.\left. -\frac{n}{J}\sum_{i=1}^{J}\left[\sigma^2_{\frac{m(k-1)}{n}}\left(\Delta W_{\frac{(k-1)m+i}{n}}\right)^2 - \left(\Delta X_{\frac{(k-1)m+i}{n}}\right)^2\right]\right|\right] \\
\leq & \sqrt{\mathrm{E}_k A^2}\sqrt{\mathrm{E}_k B^2}
\end{aligned}
$$

Let $c_i = n/m - n/J$ for $i = 1, ..., J$. $c_i = n/m$ for $i = J+1, ..., m$. Second part is square root of

$$
\begin{aligned}
& \mathrm{E}_k B^2 \\
= \; & \mathrm{E}^k \left[\frac{n}{m} \sum_{i=1}^{m} \left(\sigma^2_{\frac{m(k-1)}{n}} \left(\Delta W_{\frac{(k-1)m+i}{n}} \right)^2 - \left(\Delta X_{\frac{(k-1)m+i}{n}} \right)^2 \right) \right. \\
& \left. - \frac{n}{J} \sum_{i=1}^{J} \left(\sigma^2_{\frac{m(k-1)}{n}} \left(\Delta W_{\frac{(k-1)m+i}{n}} \right)^2 - \left(\Delta X_{\frac{(k-1)m+i}{n}} \right)^2 \right) \right]^2 \\
= \; & \mathrm{E}^k \left[\left\{ \sum_{i=1}^{m} c_i \left[\sigma^2_{\frac{m(k-1)}{n}} \left(\Delta W_{\frac{(k-1)m+i}{n}} \right)^2 - \left(\Delta X_{\frac{(k-1)m+i}{n}} \right)^2 \right] \right\}^2 \right] \\
= \; & \sum_{i=1}^{m} c_i^2 \mathrm{E}^k \left[\sigma^2_{\frac{m(k-1)}{n}} \left(\Delta W_{\frac{(k-1)m+i}{n}} \right)^2 - \left(\Delta X_{\frac{(k-1)m+i}{n}} \right)^2 \right]^2 \\
= \; & \sum_{i=1}^{m} \sum_{i'=1}^{m} c_i c_{i'} \mathrm{E}^k \left\{ \left[\sigma^2_{\frac{m(k-1)}{n}} \left(\Delta W_{\frac{(k-1)m+i}{n}} \right)^2 - \left(\Delta X_{\frac{(k-1)m+i}{n}} \right)^2 \right] \right. \\
& \left. \left[\sigma^2_{\frac{m(k-1)}{n}} \left(\Delta W_{\frac{(k-1)m+i'}{n}} \right)^2 - \left(\Delta X_{\frac{(k-1)m+i'}{n}} \right)^2 \right] \right\} \\
\leq \; & C n^{-5/2} \sum_{i=1}^{m} c_i^2 + C n^{-3} \sum_{i=1}^{m} \sum_{i'=1}^{m} |c_i| \, |c_{i'}| \\
\leq \; & C n^{-5/2} \frac{n^2}{J} + C n^{-3} n^2 \\
= \; & C n^{-1/2} J^{-1} + C n^{-1}
\end{aligned}
$$

because

$$
\begin{aligned}
& \mathrm{E}^k \left[\sigma^2_{\frac{m(k-1)}{n}} \left(\Delta W_{\frac{(k-1)m+i}{n}} \right)^2 - \left(\Delta X_{\frac{(k-1)m+i}{n}} \right)^2 \right]^2 \\
= \; & \mathrm{E}^k \left[\sigma^2_{\frac{m(k-1)}{n}} \Delta W_{\frac{(k-1)m+i}{n}} - \Delta X_{\frac{(k-1)m+i}{n}} \right]^2 \left[\sigma^2_{\frac{m(k-1)}{n}} \Delta W_{\frac{(k-1)m+i}{n}} + \Delta X_{\frac{(k-1)m+i}{n}} \right]^2 \\
\leq \; & \sqrt{\mathrm{E}^k \left[\sigma^2_{\frac{m(k-1)}{n}} \Delta W_{\frac{(k-1)m+i}{n}} - \Delta X_{\frac{(k-1)m+i}{n}} \right]^4} \sqrt{\mathrm{E}^k \left[\sigma^2_{\frac{m(k-1)}{n}} \Delta W_{\frac{(k-1)m+i}{n}} + \Delta X_{\frac{(k-1)m+i}{n}} \right]^4} \\
\leq \; & C \sqrt{\frac{1}{n^3}} \sqrt{\frac{1}{n^2}} = C n^{-5/2}.
\end{aligned}
$$

and, for $i < i'$,

$$\begin{aligned}
& \left| \mathrm{E}^k \left[\sigma^2_{\frac{m(k-1)}{n}} \left(\Delta W_{\frac{(k-1)m+i}{n}} \right)^2 - \left(\Delta X_{\frac{(k-1)m+i}{n}} \right)^2 \right] \left[\sigma^2_{\frac{m(k-1)}{n}} \left(\Delta W_{\frac{(k-1)m+i'}{n}} \right)^2 - \left(\Delta X_{\frac{(k-1)m+i'}{n}} \right)^2 \right] \right. \\
\leq \ & \mathrm{E}^k \left(\left| \sigma^2_{\frac{m(k-1)}{n}} \left(\Delta W_{\frac{(k-1)m+i}{n}} \right)^2 - \left(\Delta X_{\frac{(k-1)m+i}{n}} \right)^2 \right| \right. \\
& \left. \mathrm{E} \left[\left| \sigma^2_{\frac{m(k-1)}{n}} \left(\Delta W_{\frac{(k-1)m+i'}{n}} \right)^2 - \left(\Delta X_{\frac{(k-1)m+i'}{n}} \right)^2 \right| \middle| \mathcal{F}_{\frac{(k-1)m+i}{n}} \right] \right) \\
\leq \ & Cn^{-3/2} \mathrm{E}^k \left[\left| \sigma^2_{\frac{m(k-1)}{n}} \left(\Delta W_{\frac{(k-1)m+i}{n}} \right)^2 - \left(\Delta X_{\frac{(k-1)m+i}{n}} \right)^2 \right| \right] \\
\leq \ & Cn^{-3}.
\end{aligned}$$

First part is square root of

$$\begin{aligned}
\mathrm{E}_k A^2 &= \mathrm{E}_k \left(\widehat{\alpha}_k^{long} - \widehat{\alpha}_k^{short} + \widehat{\theta}_k^{long} - \widehat{\theta}_k^{short} \right)^2 \\
&= \mathrm{E}^k \left[\left\{ \sum_{i=1}^{m} c_i \left[\sigma^2_{\frac{m(k-1)}{n}} \left(\Delta W_{\frac{(k-1)m+i}{n}} \right)^2 + \left(\Delta X_{\frac{(k-1)m+i}{n}} \right)^2 \right] \right\}^2 \right] \\
&\leq C.
\end{aligned}$$

Combining both A and B terms, we obtain

$$\mathrm{E} \left[\left| \gamma_k^{DISCR} - \gamma_k \right| \middle| \mathcal{F}_{\frac{k-1}{K}} \right] \leq Cn^{-1/4} J^{-1/2} + Cn^{-1/2},$$

from which second step

$$\left| \mathcal{E}\left(\widehat{V}\right)^{DISCR} - \mathcal{E}\left(\widehat{V}\right) \right| \leq \frac{J}{K} \sum_{k=1}^{K} \mathrm{E} \left[\left| \gamma_k^{DISCR} - \gamma_k \right| \middle| \mathcal{F}_{\frac{k-1}{K}} \right] \leq CJn^{-1/4} J^{-1/2} + CJn^{-1/2} \xrightarrow{p} 0$$

follows, provided $J^2/n \to 0$, which is implied by $mJ^2 n^{-1} \to 0$.

Now we prove the third step.

$$\begin{aligned}
& \mathrm{E}\left[\gamma_k^{DISCR}\middle|\mathcal{F}_{\frac{k-1}{K}}\right] \\
= \; & \sigma^4_{\frac{m(k-1)}{n}} \mathrm{E}\left[\left\{\frac{n}{J}\sum_{i=1}^{J}\left(W_{\frac{(k-1)m+i}{n}} - W_{\frac{(k-1)m+i-1}{n}}\right)^2 - \frac{n}{m}\sum_{i=1}^{m}\left(W_{\frac{(k-1)m+i}{n}} - W_{\frac{(k-1)m+i-1}{n}}\right)^2\right\}^2\right] \\
= \; & \sigma^4_{\frac{m(k-1)}{n}}\frac{2}{J} - \sigma^4_{\frac{m(k-1)}{n}}\frac{2}{m}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathcal{E}\left(\widehat{V}\right) &= \frac{J}{K}\sum_{k=1}^{K}\mathrm{E}\left[\gamma_k\middle|\mathcal{F}_{\frac{k-1}{K}}\right] \\
&= \frac{J}{K}\sum_{k=1}^{K}\sigma^4_{\frac{m(k-1)}{n}}\frac{2}{J} - \frac{J}{K}\sum_{k=1}^{K}\sigma^4_{\frac{m(k-1)}{n}}\frac{2}{m} \\
&= \widehat{V}_{sub}^{DISCR} - O_p\left(\frac{J}{m}\right).
\end{aligned}$$

This proves consistency of the subsampling method for RV, provided $mJ^2n^{-1} \to 0$ and σ satisfies A1. ∎

B.1.3 Proof of Proposition 3

Proposition 3 is proved for the special case $Q = m$. The general Q case follows by the same steps, but the notation is more involved. Denote $K = \lfloor n/m \rfloor$ and $\Delta_\delta X_t = X_t - X_{t-\delta}$.

Introduce the same notation as in Proposition 2.

$$\begin{aligned}
V^{DISCR} &= \frac{m}{n}\sum_{k=1}^{K} 2\sigma^4_{\frac{k-1}{K}} & \mathcal{E}\left(\widehat{V}\right)^{DISCR} &= \frac{m}{n}\sum_{k=1}^{K}\left[\mathrm{E}\gamma_k^{DISCR}\middle|\mathcal{F}_{\frac{k-1}{K}}\right] \\
\mathcal{E}\left(\widehat{V}\right) &= \frac{m}{n}\sum_{k=1}^{K}\mathrm{E}\left[\gamma_k\middle|\mathcal{F}_{\frac{k-1}{K}}\right] & \widehat{\alpha}_k^{slow} &= \sigma^2_{\frac{m(k-1)}{n}}\left(\Delta_{\frac{m}{n}}W_{\frac{mk}{n}}\right)^2 \\
\gamma_k &= \left(\widehat{\theta}_k^{slow} - \widehat{\theta}_k^{fast}\right)^2 & \widehat{\alpha}_k^{fast} &= \sigma^2_{\frac{m(k-1)}{n}}\sum_{i=1}^{m}\left(\Delta_{\frac{1}{n}}W_{\frac{i+m(k-1)}{n}}\right)^2 \\
\gamma_k^{DISCR} &= \left(\widehat{\alpha}_k^{slow} - \widehat{\alpha}_k^{fast}\right)^2
\end{aligned}$$

Also, denote $\mathrm{E}\left[\gamma_k \left|\mathcal{F}_{\frac{k-1}{K}}\right.\right]$ by $\mathrm{E}^n_{k-1}[\gamma_k]$. We want to show

$$\widehat{V}_a \xrightarrow{p} V = 2\int_0^1 \sigma_u^4 du$$

where

$$\widehat{V}_a = \frac{n}{m}\sum_{k=1}^{K}\left(\widehat{\theta}_k^{slow} - \widehat{\theta}_k^{fast}\right)^2 = \frac{n}{m}\sum_{k=1}^{K}\left\{\left(\Delta_{\frac{m}{n}} X_{\frac{mk}{n}}\right)^2 - \sum_{i=1}^{m}\left(\Delta_{\frac{1}{n}} X_{\frac{i+m(k-1)}{n}}\right)^2\right\}^2.$$

First, by Riemann integrability,

$$V^{DISCR} \xrightarrow{p} V = 2\int_0^1 \sigma_u^4 du. \tag{B.3}$$

To prove Proposition 3, use the following three steps. Prove $\widehat{V} - \mathcal{E}\left(\widehat{V}\right) \xrightarrow{p} 0$, then $\mathcal{E}\left(\widehat{V}\right)^{DISCR} - \mathcal{E}\left(\widehat{V}\right) \xrightarrow{p} 0$, and finally $\mathcal{E}\left(\widehat{V}\right)^{DISCR} - V^{DISCR} \xrightarrow{p} 0$.

The first step is to show

$$\widehat{V} - \mathcal{E}\left(\widehat{V}\right) = K\sum_{k=1}^{K}\left(\gamma_k - E\left[\gamma_k \left|\mathcal{F}_{\frac{k-1}{K}}\right.\right]\right) \xrightarrow{p} 0.$$

By Lenglart's inequality (see e.g. Podolskij 2006), it is sufficient to show that

$$\sum_{k=1}^{K} E\left[\left|K\gamma_k\right|^2 \left|\mathcal{F}_{\frac{k-1}{K}}\right.\right] \xrightarrow{p} 0.$$

Notice that, by Burkholder-Davis-Gundy inequality, Cauchy-Schwarz inequality, and uniform boundedness of σ,

$$
\begin{aligned}
& \mathrm{E}^n_{k-1}\left[\left(\widehat{\theta}_k^{fast}\right)^4\right] \\
\leq\ & \sum_{i'''=1}^{m}\sum_{i''=1}^{m}\sum_{i'=1}^{m}\sum_{i=1}^{m} \sqrt[4]{\mathrm{E}^n_{k-1}\left[\left(\Delta_{\frac{1}{n}} X_{\frac{i+m(k-1)}{n}}\right)^8\right]} \sqrt[4]{\mathrm{E}^n_{k-1}\left[\left(\Delta_{\frac{1}{n}} X_{\frac{i'+m(k-1)}{n}}\right)^8\right]} \times \\
& \sqrt[4]{\mathrm{E}^n_{k-1}\left[\left(\Delta_{\frac{1}{n}} X_{\frac{i''+m(k-1)}{n}}\right)^8\right]} \sqrt[4]{\mathrm{E}^n_{k-1}\left[\left(\Delta_{\frac{1}{n}} X_{\frac{i'''+m(k-1)}{n}}\right)^8\right]} \\
\leq\ & C\frac{m^4}{n^4} = C\frac{1}{K^4}
\end{aligned}
$$

for some constant C, which does not depend on any of the above parameters. Hence, and by similarity,

$$
\begin{aligned}
\mathrm{E}^n_{k-1}\left[\left(\widehat{\theta}_k^{fast}\right)^4\right] &\leq \frac{C}{K^4},\ \mathrm{E}^n_{k-1}\left[\left(\widehat{\theta}_k^{fast}\right)^3\right] \leq \frac{C}{K^4},\ \mathrm{E}^n_{k-1}\left[\left(\widehat{\theta}_k^{fast}\right)^2\right] \leq \frac{C}{K^2}, \qquad \text{(B.4)} \\
\mathrm{E}^n_{k-1}\left[\left(\widehat{\theta}_k^{slow}\right)^4\right] &\leq \frac{C}{K^4},\ \mathrm{E}^n_{k-1}\left[\left(\widehat{\theta}_k^{slow}\right)^3\right] \leq \frac{C}{K^3},\ \mathrm{E}^n_{k-1}\left[\left(\widehat{\theta}_k^{slow}\right)^2\right] \leq \frac{C}{K^2}.
\end{aligned}
$$

From here,

$$
\mathrm{E}^n_{k-1}\left[\gamma_k^2\right] = \mathrm{E}^n_{k-1}\left[\left(\widehat{\theta}_k^{fast} - \widehat{\theta}_k^{slow}\right)^4\right] \leq C\frac{1}{K^4}
$$

and

$$
\sum_{k=1}^{K} \mathrm{E}\left[\left|K\gamma_k\right|^2 \middle| \mathcal{F}_{\frac{k-1}{K}}\right] \leq C\frac{1}{K} = o(1).
$$

The second step is to show

$$
\mathcal{E}\left(\widehat{V}\right)^{DISCR} - \mathcal{E}\left(\widehat{V}\right) = K\sum_{k=1}^{K} \mathrm{E}\left[\gamma_k^{DISCR} - \gamma_k \middle| \mathcal{F}_{\frac{k-1}{K}}\right] \xrightarrow{p} 0.
$$

It is sufficient to show

$$
K\sum_{k=1}^{K} \mathrm{E}\left[\left|\gamma_k^{DISCR} - \gamma_k\right|\right] \rightarrow 0.
$$

Write

$$
\begin{aligned}
K \sum_{k=1}^{K} \mathrm{E}\left[\left|\gamma_k^{DISCR} - \gamma_k\right|\right] &= K \sum_{k=1}^{K} \mathrm{E}\left[\left|\widehat{\alpha}_k^{fast} - \widehat{\alpha}_k^{slow} + \widehat{\theta}_k^{fast} - \widehat{\theta}_k^{slow}\right|\right. \\
&\quad \left.\left|\left\{\widehat{\alpha}_k^{fast} - \widehat{\theta}_k^{fast}\right\} - \left\{\widehat{\alpha}_k^{slow} - \widehat{\theta}_k^{slow}\right\}\right|\right] \\
&\equiv A + B
\end{aligned}
$$

As to the first term, we have

$$
\begin{aligned}
A &= K\sum_{k=1}^{K} \mathrm{E}\left[\left|\widehat{\alpha}_k^{fast} - \widehat{\alpha}_k^{slow} + \widehat{\theta}_k^{fast} - \widehat{\theta}_k^{slow}\right|\left|\widehat{\alpha}_k^{fast} - \widehat{\theta}_k^{fast}\right|\right] \\
&\leq K\sum_{k=1}^{K}\left\{\mathrm{E}\left[\widehat{\alpha}_k^{fast} - \widehat{\alpha}_k^{slow} + \widehat{\theta}_k^{fast} - \widehat{\theta}_k^{slow}\right]^2\right\}^{1/2}\left\{\mathrm{E}\left[\widehat{\alpha}_k^{fast} - \widehat{\theta}_k^{fast}\right]^2\right\}^{1/2} \\
&\leq C\sum_{k=1}^{K}\left\{\mathrm{E}\left[\widehat{\alpha}_k^{fast} - \widehat{\theta}_k^{fast}\right]^2\right\}^{1/2} \\
&= C\sum_{k=1}^{K}\left\{\mathrm{E}\left[\sum_{i=1}^{m}\left\{\sigma_{\frac{m(k-1)}{n}}\Delta_{\frac{1}{n}}W_{\frac{i+m(k-1)}{n}} - \Delta_{\frac{1}{n}}X_{\frac{i+m(k-1)}{n}}\right\}\right.\right. \\
&\quad \left.\left.\left\{\sigma_{\frac{m(k-1)}{n}}\Delta_{\frac{1}{n}}W_{\frac{i+m(k-1)}{n}} + \Delta_{\frac{1}{n}}X_{\frac{i+m(k-1)}{n}}\right\}\right]^2\right\}^{1/2} \\
&\leq C\sum_{k=1}^{K}\left\{\sum_{i'=1}^{m}\sum_{i=1}^{m}\sqrt[4]{\mathrm{E}\left[\left\{\sigma_{\frac{m(k-1)}{n}}\Delta_{\frac{1}{n}}W_{\frac{i+m(k-1)}{n}} - \Delta_{\frac{1}{n}}X_{\frac{i+m(k-1)}{n}}\right\}^4\right]}\times\right. \\
&\quad \sqrt[4]{\mathrm{E}\left[\left\{\sigma_{\frac{m(k-1)}{n}}\Delta_{\frac{1}{n}}W_{\frac{i+m(k-1)}{n}} + \Delta_{\frac{1}{n}}X_{\frac{i+m(k-1)}{n}}\right\}^4\right]}\times \\
&\quad \sqrt[4]{\mathrm{E}\left[\left\{\sigma_{\frac{m(k-1)}{n}}\Delta_{\frac{1}{n}}W_{\frac{i'+m(k-1)}{n}} - \Delta_{\frac{1}{n}}X_{\frac{i'+m(k-1)}{n}}\right\}^4\right]} \\
&\quad \left.\sqrt[4]{\mathrm{E}\left[\left\{\sigma_{\frac{m(k-1)}{n}}\Delta_{\frac{1}{n}}W_{\frac{i'+m(k-1)}{n}} + \Delta_{\frac{1}{n}}X_{\frac{i'+m(k-1)}{n}}\right\}^4\right]}\right\}^{1/2} \\
&\leq \frac{C}{\sqrt{n}}\sum_{k=1}^{K}\sum_{i=1}^{m}\sqrt[4]{\mathrm{E}\left[\left\{\sigma_{\frac{m(k-1)}{n}}\Delta_{\frac{1}{n}}W_{\frac{i+m(k-1)}{n}} - \Delta_{\frac{1}{n}}X_{\frac{i+m(k-1)}{n}}\right\}^4\right]} \\
&\leq \frac{C}{\sqrt{n}}\sum_{k=1}^{K}\sum_{i=1}^{m}\left\{\mathrm{E}\left[\int_{\frac{i-1+m(k-1)}{n}}^{\frac{i+m(k-1)}{n}}\left(\sigma_{\frac{m(k-1)}{n}} - \sigma_u\right)^2 du\right]^2\right\}^{1/4} \\
&= \frac{C}{\sqrt{n}}\sum_{k=1}^{K}\sum_{i=1}^{m}\left\{\mathrm{E}\left[\int_{\frac{i-1+m(k-1)}{n}}^{\frac{i+m(k-1)}{n}}\left(\sigma_{\frac{\lfloor Ku\rfloor}{K}} - \sigma_u\right)^2 du\right]^2\right\}^{1/4}
\end{aligned}
$$

In above, to obtain second inequality, we used (B.4). To obtain the fourth inequality, we used

$$
\begin{aligned}
& \mathrm{E}\left[\left\{\sigma_{\frac{m(k-1)}{n}} \Delta_{\frac{1}{n}} W_{\frac{i+m(k-1)}{n}} + \Delta_{\frac{1}{n}} X_{\frac{i+m(k-1)}{n}}\right\}^4\right] \\
\leq \ & C\mathrm{E}\left[\left\{\int_{\frac{i-1+m(k-1)}{n}}^{\frac{i+m(k-1)}{n}} \left(\sigma_{\frac{m(k-1)}{n}} + \sigma_u\right)^2 du\right\}^2\right] \\
\leq \ & \frac{C}{n^2},
\end{aligned}
$$

which follows by Burkholder-Davis-Gundy inequality. To proceed with term A, we use the arguments along the lines of the proof of Lemma 1 of Barndorff-Nielsen (2001). For every i and k, there exists a constant $c_{i,k}$ s.t.

$$
\inf_{\frac{i-1+m(k-1)}{n} \leq u \leq \frac{i+m(k-1)}{n}} \left(\sigma_{\frac{\lfloor Ku \rfloor}{K}} - \sigma_u\right)^2 \leq c_{i,k} \leq \sup_{\frac{i-1+m(k-1)}{n} \leq u \leq \frac{i+m(k-1)}{n}} \left(\sigma_{\frac{\lfloor Ku \rfloor}{K}} - \sigma_u\right)^2
$$

and

$$
\int_{\frac{i-1+m(k-1)}{n}}^{\frac{i+m(k-1)}{n}} \left(\sigma_{\frac{\lfloor Ku \rfloor}{K}} - \sigma_u\right)^2 du = c_{i,k}\frac{1}{n}.
$$

Notice that

$$
\sup_{i,k} c_{i,k} \to 0
$$

by right-continuity and boundedness of σ. Then,

$$
\begin{aligned}
A \ & \leq \ \frac{C}{\sqrt{n}} \sum_{k=1}^{K} \sum_{i=1}^{m} \left\{\mathrm{E}\left[\int_{\frac{i-1+m(k-1)}{n}}^{\frac{i+m(k-1)}{n}} \left(\sigma_{\frac{\lfloor Ku \rfloor}{K}} - \sigma_u\right)^2 du\right]^2\right\}^{1/4} \\
& = \ \frac{C}{\sqrt{n}} \sum_{k=1}^{K} \sum_{i=1}^{m} \left\{\mathrm{E}\left[c_{i,k}\frac{1}{n}\right]^2\right\}^{1/4} = C \sum_{k=1}^{K} \sum_{i=1}^{m} \sqrt[4]{\mathrm{E}c_{i,k}^2}\frac{1}{n} \to 0
\end{aligned}
$$

by Monotone Convergence Theorem. $B \to 0$ is proved using exactly the same steps. This proves the second step.

The final step is to show

$$\mathcal{E}\left(\widehat{V}\right)^{DISCR} - V^{DISCR} \xrightarrow{p} 0.$$

We have

$$\begin{aligned}
& \mathrm{E}\left[\gamma_k^{DISCR} \middle| \mathcal{F}_{\frac{k-1}{K}}\right] \\
= \; & \sigma^4_{\frac{m(k-1)}{n}} \mathrm{E}\left[\left\{\left(\Delta_{\frac{m}{n}} W_{\frac{mk}{n}}\right)^2 - \sum_{i=1}^{m}\left(\Delta_{\frac{1}{n}} W_{\frac{i+m(k-1)}{n}}\right)^2\right\}^2\right] \\
= \; & \frac{2}{K^2}\sigma^4_{\frac{m(k-1)}{n}} + o_p\left(\frac{1}{K^2}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathcal{E}\left(\widehat{V}\right)^{DISCR} &= K\sum_{k=1}^{K} \mathrm{E}\left[\gamma_k^{DISCR} \middle| \mathcal{F}_{\frac{k-1}{K}}\right] = K\sum_{k=1}^{K}\left(\frac{2}{K^2}\sigma^4_{\frac{m(k-1)}{n}} + o_p\left(\frac{1}{K^2}\right)\right) \\
&= \sum_{k=1}^{K}\frac{2}{K^1}\sigma^4_{\frac{m(k-1)}{n}} + o_p(1) = V^{DISCR} + o_p(1).
\end{aligned}$$

The result follows immediately. ∎

B.1.4 Proof of Theorem 2.3.1

To simplify the proof, assume $\mu \equiv 0$. Below arguments can easily be extended to nonzero drift. Recall that Theorem 2.3.1 assumes $\{\sigma\} \perp \{W\}$. Therefore, all arguments are done conditional on volatility. This Theorem is proved by validating the conditions of Theorem 2.4.1 for the Two Time Scale estimator of Aït-Sahalia et al. (2006a).

We first introduce some additional notation. For some arguments w and q, define

$$\overline{q}_w = \frac{q - w + 1}{w}.$$

This is the average numbers of observations in a (Infill Price type) subsample if total number of observations is q and there are w subsamples.

It is convenient to decompose the variance into signal and noise part, $V_l^{short} = V_l^{signal} + V_l^{noise}$ where

$$\begin{aligned} V_l^{signal} &= \frac{4}{3}c \int_{(l-1)m/n}^{[(l-1)m+J]/n} \sigma_u^4 du \\ V_l^{noise} &= 8c^{-2}\frac{J}{n}\text{Var}(\epsilon)^2 + 16\frac{J}{n}c^{-2} \lim_{n\to\infty} \sum_{i=1}^{n} \text{Cov}\left(\epsilon_0, \epsilon_{i/n}\right)^2. \end{aligned}$$

We now prove Assumption A6(i) is satisfied. For that, we decompose $\zeta_l^{(n)}$ into signal and noise components. For the signal part, we show that arguments of Aït-Sahalia et al. (2006a) carry over to the subsample. For the noise part, their arguments apply directly.

The decomposition is

$$\begin{aligned} & \mathrm{E}\left(\frac{n}{J}\left(n^{1/3}\left(\widehat{\theta}_l^{short} - \theta_l^{short}\right)^2 - V_l^{short}\right)\right) \\ =\ & \frac{n}{J}\left(n^{1/3}\mathrm{E}\left[\left([Y,Y]_l^{(G_1)} - \frac{\overline{J}_{G_1}}{\overline{J}_{G_2}}[Y,Y]_l^{(G_2)} - \theta_l^{short}\right)^2\right] - V_l^{short}\right) \\ =\ & \frac{n}{J}\left(n^{1/3}\mathrm{E}\left[\left([X,X]_l^{(G_1)} - \theta_l^{short} + [\epsilon,\epsilon]_l^{(G_1)} - \frac{\overline{J}_{G_1}}{\overline{J}_{G_2}}[\epsilon,\epsilon]_l^{(G_2)}\right)^2\right] - V_l^{short}\right) + R_1 \\ =\ & \frac{n^{4/3}}{J}\mathrm{E}\left([X,X]_l^{(G_1)} - \theta_l^{short}\right)^2 + \frac{n^{4/3}}{J}\mathrm{E}\left([\epsilon,\epsilon]_l^{(G_1)} - \frac{\overline{J}_{G_1}}{\overline{J}_{G_2}}[\epsilon,\epsilon]_l^{(G_2)}\right)^2 \\ & -\frac{n}{J}V_l^{short} + R_1 + R_2. \end{aligned}$$

As a first step, we show negligibility of the signal part, i.e.,

$$\frac{n^{4/3}}{J}\mathrm{E}\left([X,X]_l^{(G_1)} - \theta_l^{short}\right)^2 - \frac{n}{J}V_l^{signal} = o(1). \tag{B.5}$$

For this, we adapt the arguments of Zhang et al. (2005) to the subsample. We have

$$[X,X]_l^{(G_1)} = [X,X]_l^{(1)} + S_l + R_3$$

where

$$S_l = 2\sum_{i=1}^{J-1} \left(\Delta X_{(l-1)m/n+i/n}\right) \sum_{j=1}^{G_1 \wedge i} \left(1 - \frac{j}{G_1}\right) \left(\Delta X_{(l-1)m/n+(i-j)/n}\right)$$

where $\Delta X_{i/n} = X_{i/n} - X_{(i-1)/n}$. R_3 arises due to the end effects, see Zhang et al. (2005), p.1410., and it satisfies

$$\mathrm{E}\left(R_3\right) \leq C\frac{G_1}{n},\ \mathrm{E}\left(R_3^2\right) \leq C\frac{G_1^2}{n^2}$$

By (B.6) and (B.7), we have $\mathrm{E}\left(S_l^2\right) \leq Jn^{-4/3}$ and $n^{1/3}\mathrm{E}\left([X,X]_l^{(1)} - \theta_l^{short}\right)^2 \leq Cn^{-5/3}$. Therefore, to prove (B.5), it is sufficient to show

$$\frac{n^{4/3}}{J}\mathrm{E}\left([X,X]_l^{(1)} + S_l - \theta_l^{short}\right)^2 - \frac{n}{J}V_l^{signal} = o\left(1\right).$$

We have

$$\begin{aligned}
& \mathrm{E}\left(S_l^2\right) \qquad\qquad \text{(B.6)} \\
= \ & \mathrm{E}\left[\left(2\sum_{i=1}^{J-1}\left(\Delta X_{[(l-1)m+i]/n}\right)\sum_{j=1}^{G_1\wedge i}\left(1-\frac{j}{G_1}\right)\left(\Delta X_{[(l-1)m+i-j]/n}\right)\right)^2\right] \\
= \ & 4\sum_{i=1}^{J-1}\int_{[(l-1)m+i-1]/n}^{[(l-1)m+i]/n}\sigma_u^2du\sum_{j=1}^{G_1\wedge i}\left(1-\frac{j}{G_1}\right)^2\int_{((l-1)m+i-j-1)/n}^{((l-1)m+i-j)/n}\sigma_u^2du \\
= \ & 4\sum_{i=G_1}^{J-1}\int_{[(l-1)m+i-1]/n}^{[(l-1)m+i]/n}\sigma_u^2du\sum_{j=1}^{G_1}\left(1-\frac{j}{G_1}\right)^2\int_{((l-1)m+i-j-1)/n}^{((l-1)m+i-j)/n}\sigma_u^2du + o\left(\frac{J}{n^{4/3}}\right) \\
= \ & 4\sum_{i=G_1}^{J-1}\frac{\sigma^2_{[(l-1)m+i]/n}}{n}\sum_{j=1}^{G_1}\left(1-\frac{j}{G_1}\right)^2\frac{\sigma^2_{[(l-1)m+i]/n}}{n} + R_4 + R_5 + o\left(\frac{J}{n^{4/3}}\right) \\
= \ & \frac{4}{3}\frac{G_1}{n^2}\sum_{i=G_1}^{J-1}\sigma^4_{[(l-1)m+i]/n} + R_4 + R_5 + o\left(\frac{J}{n^{4/3}}\right) \\
= \ & \frac{4}{3}\frac{c}{n^{1/3}}\int_0^{J/n}\sigma_u^4du + R_4 + R_5 + o\left(\frac{J}{n^{4/3}}\right)
\end{aligned}$$

where we use $G_1 = cn^{2/3}$.

The remainder terms R_4 and R_5 have expressions as below, and they are of smaller order than $Jn^{-4/3}$ by (B.1),

$$\begin{aligned}
&\frac{n^{4/3}}{J} R_4 \\
= &\frac{n^{4/3}}{J} 4 \sum_{i=G_1}^{J-1} \frac{\sigma^2_{[(l-1)m+i]/n}}{n} \sum_{j=1}^{G_1} \left(1 - \frac{j}{G_1}\right)^2 \int_{((l-1)m+i-j-1)/n}^{((l-1)m+i-j)/n} \left(\sigma_u^2 - \sigma_{i/n}^2\right) du \\
\leq &C \frac{n^{4/3}}{J} 4 \sum_{i=G_1}^{J-1} \frac{\sigma^2_{[(l-1)m+i]/n}}{n} \sum_{j=1}^{G_1} \left(1 - \frac{j}{G_1}\right)^2 \frac{1}{n} \sqrt{\frac{G_1}{n}} \leq C \frac{n^{4/3}}{J} \frac{1}{n^2} \sqrt{\frac{G_1}{n}} J G_1 \to 0,
\end{aligned}$$

$$\begin{aligned}
&\frac{n^{4/3}}{J} R_5 \\
= &\frac{n^{4/3}}{J} \frac{n^{4/3}}{J} 4 \sum_{i=G_1}^{J-1} \int_{[(l-1)m+i-1]/n}^{[(l-1)m+i]/n} \left(\sigma_u^2 - \sigma_{i/n}^2\right) du \sum_{j=1}^{G_1} \left(1 - \frac{j}{G_1}\right)^2 \int_{((l-1)m+i-j-1)/n}^{((l-1)m+i-j)/n} \sigma_u^2 du \\
\leq &C \frac{n^{4/3}}{J} 4 \sum_{i=G_1}^{J-1} \frac{1}{n^{3/2}} \sum_{j=1}^{G_1} \left(1 - \frac{j}{G_1}\right)^2 \frac{1}{n} \leq C \frac{n^{4/3}}{J} \frac{1}{n^{3/2}} \frac{1}{n} J G_1 \to 0.
\end{aligned}$$

Therefore,

$$\frac{n^{4/3}}{J} \mathrm{E}\left(S_l^2\right) - \frac{n}{J} V_l^{signal} = o(1).$$

Finally,

$$\frac{n^{4/3}}{J} \mathrm{E}\left([X,X]_l^{(1)} - \theta_l^{short}\right)^2 = \frac{n^{4/3}}{J} \mathrm{Var}\left([X,X]_l^{(1)}\right) \leq C \frac{n^{4/3}}{J} \frac{J}{n^2} = o(1). \qquad \text{(B.7)}$$

This concludes the proof of (B.5). Next, we turn to the noise part and prove

$$n^{1/3} \mathrm{E}\left([\epsilon,\epsilon]_l^{(G_1)} - \frac{\bar{J}_{G_1}}{\bar{J}_{G_2}} [\epsilon,\epsilon]_l^{(G_2)}\right)^2 - V_l^{short} = o\left(\frac{J}{n}\right). \qquad \text{(B.8)}$$

In this case, Proposition 1 of Aït-Sahalia et al. (2006a) can be applied directly, with J instead of n (this is the number of observations used above) to obtain

$$\begin{aligned}
&\frac{G_1}{\sqrt{J}}\left([\epsilon,\epsilon]_l^{(G_1)}-\frac{\overline{J}_{G_1}}{\overline{J}_{G_2}}[\epsilon,\epsilon]_l^{(G_2)}\right) \\
&\Rightarrow N\left(0, 8\mathrm{Var}(\epsilon)^2+16\lim_{n\to\infty}\sum_{i=1}^{n}\mathrm{Cov}\left(\epsilon_0,\epsilon_{i/n}\right)^2\right)
\end{aligned}$$

$$\begin{aligned}
\frac{G_1^2}{J}\mathrm{E}\left([\epsilon,\epsilon]_l^{(G_1)}-\frac{\overline{J}_{G_1}}{\overline{J}_{G_2}}[\epsilon,\epsilon]_l^{(G_2)}\right)^2 &\to 8\mathrm{Var}(\epsilon)^2+16\lim_{n\to\infty}\sum_{i=1}^{n}\mathrm{Cov}\left(\epsilon_0,\epsilon_{i/n}\right)^2 \\
\frac{n^{4/3}}{J}\mathrm{E}\left([\epsilon,\epsilon]_l^{(G_1)}-\frac{\overline{J}_{G_1}}{\overline{J}_{G_2}}[\epsilon,\epsilon]_l^{(G_2)}\right)^2-\frac{n}{J}V^{noise} &\to 0,
\end{aligned}$$

which immediately gives (B.8).

The final step is to show $R_1+R_2=o(1)$. We have

$$\begin{aligned}
R_1 &= \frac{n}{J}n^{1/3}\mathrm{E}\left[\left(\frac{\overline{J}_{G_1}}{\overline{J}_{G_2}}[X,X]_l^{(G_2)}\right)^2\right]+\frac{n}{J}n^{1/3}\mathrm{E}\left[\left(2[X,\epsilon]_l^{(G_1)}\right)^2\right] \\
&\quad+\left(\frac{n}{J}\right)^{1/3}\mathrm{E}\left[\left(\frac{\overline{J}_{G_1}}{\overline{J}_{G_2}}[X,\epsilon]_l^{(G_2)}\right)^2\right]+R_1'
\end{aligned}$$

where, for $i=1,2$,

$$[X,\epsilon]_l^{(G_i)}=\frac{1}{G_1}\sum_{i=1}^{n-G_1}\left(X_{(i+G_1)/n}-X_{i/n}\right)\left(\epsilon_{(i+G_1)/n}-\epsilon_{i/n}\right).$$

First term is $o(1)$ because $[X,X]_l^{(G_2)}\xrightarrow{p}\theta_l^{short}$ by substituting G_2 for G_1 in (B.5). Second and third terms are of smaller order than m/n by proof of Lemma 1 of Aït-Sahalia et al. (2006a), which implies, for $i=1,2$,

$$\mathrm{E}\left(\left([X,\epsilon]_l^{(G_i)}\right)^2|X\right)\leq C\frac{1}{G_i^2}[X,X]_l^{(G_i)}.$$

The final terms R' and R_2 contain cross terms that are negligible by Cauchy-Schwarz inequality. This concludes verification of the assumption A6(i).

Assumption A6(ii) can be verified using straightforward calculations, by using the mixing property of the noise ϵ, as well as $L^{4+\delta}$ boundedness of ϵ for some δ, and finally

$$\mathrm{E}\left(|X_t - X_{t-s}|^q\right) \leq C_q s^{q/2},$$

for some constant C_q depending on q only, for all $q > 0$ and all s. Above follows from Burkholder-Davis-Gundy inequality.

Assumption A5 is immediate due to $\{\mu\},\{\sigma\} \perp \{W\}$ assumption. This implies that conditional on volatility and drift path, returns are independent over non-overlapping intervals. Hence, strong mixing of $\zeta_l^{(n)}$ follows from strong mixing of the ϵ, which in turn holds by assumption. This concludes the proof of Theorem 2.3.1. ▮

B.1.5 Proof of Lemma 2.3.2

Most of the proof of the asymptotic distribution of TSRV estimator of Aït-Sahalia et al. (2006a) remains valid under the assumptions of Lemma 2.3.2. The noise component of the asymptotic distribution arises from the asymptotic distribution of

$$-2\frac{1}{\sqrt{n}}\sum_{i=0}^{n-G_1}\epsilon_{i/n}\epsilon_{(i+G_1)/n} + 2\frac{1}{\sqrt{n}}\sum_{i=0}^{n-G_2}\epsilon_{i/n}\epsilon_{(i+G_2)/n},$$

see page 26 of Aït-Sahalia et al. (2006a). Given that $G_1/G_2 \to 0$ and

$$\omega\left(\frac{i+G_1}{n}\right) - \omega\left(\frac{i}{n}\right) \leq C\frac{G_1}{n}$$

due to differentiability of ω, the desired result follows. ▮

B.1.6 Proof of Theorem 2.4.1

Assume n is divisible by m by simplicity. As a first step, we prove

$$G^{(n)} = \frac{mJ}{n^2} \sum_{l=1}^{n/m} \tau_n^2 \left(\frac{n}{J} \widehat{\theta}_l^{short} - \frac{n}{J} \theta_l^{short} \right)^2 \xrightarrow{p} V \tag{B.9}$$

For any two subsamples l and l' s.t. $l \neq l'$, $\zeta_l^{(n)}$ has no common returns with $\zeta_{l'}^{(n)}$. Therefore, $\zeta_l^{(n)}$ is strong mixing because $R^{(n)}$ is. Moreover, if we define

$$\psi_i^{(n)} = \zeta_l^{(n)} - \mathrm{E}\left(\zeta_l^{(n)}\right),$$

it is also strong mixing. Therefore, under A6, $\psi_i^{(n)}$ is a uniformly integrable L^1-mixingale as defined in Andrews (1998), to which we can apply Theorem 2 of Andrews (1998) to obtain

$$\frac{m}{n} \sum_{l=1}^{n/m} \psi_i^{(n)} = \frac{m}{n} \sum_{l=1}^{n/m} \left[\zeta_l^{(n)} - \mathrm{E}\left(\zeta_l^{(n)}\right) \right] \xrightarrow{p} 0.$$

By A4, we have

$$\begin{aligned} \frac{m}{n} \sum_{l=1}^{n/m} \zeta_l^{(n)} &= \frac{m}{n} \sum_{l=1}^{n/m} \frac{n}{J} \left[\tau_n^2 \left(\widehat{\theta}_l^{short} - \theta_l^{short} \right)^2 - V_l^{short} \right] \xrightarrow{p} 0 \\ & \frac{m}{n} \sum_{l=1}^{n/m} \frac{n}{J} \tau_n^2 \left(\widehat{\theta}_l^{short} - \theta_l^{short} \right)^2 - \frac{m}{n} \sum_{l=1}^{n/m} \frac{n}{J} V_l^{short} \xrightarrow{p} 0 \\ & \frac{m}{n} \sum_{l=1}^{n/m} \frac{n}{J} \tau_n^2 \left(\widehat{\theta}_l^{short} - \theta_l^{short} \right)^2 \xrightarrow{p} V \end{aligned}$$

and so (B.9) follows.

In a second step, we prove that $G^{(n)} - \widehat{V} \xrightarrow{p} 0$.

$$
\begin{aligned}
& \widehat{V} - G^{(n)} & \text{(B.10)}\\
= \; & \frac{Jm\tau_n^2}{n^2}\sum_{l=1}^{K}\left(\frac{n}{J}\widehat{\theta}_l^{short} - \frac{n}{m}\widehat{\theta}_l^{long}\right)^2 - \frac{Jm\tau_n^2}{n^2}\sum_{l=1}^{n/m}\tau_n^2\left(\frac{n}{J}\widehat{\theta}_l^{short} - \frac{n}{J}\theta_l^{short}\right)^2 & \text{(B.11)}\\
= \; & \frac{Jm\tau_n^2}{n^2}\sum_{l=1}^{K}\left(\frac{n}{m}\widehat{\theta}_l^{long} - \frac{n}{m}\theta_l^{long}\right)^2 + \frac{Jm\tau_n^2}{n^2}\sum_{l=1}^{K}\left(\frac{n}{m}\theta_l^{long} - \frac{n}{J}\theta_l^{short}\right)^2 \\
& +2\frac{Jm\tau_n^2}{n^2}\sum_{l=1}^{K}\left(\frac{n}{m}\widehat{\theta}_l^{long} - \frac{n}{m}\theta_l^{long}\right)\left(\frac{n}{m}\theta_l^{long} - \frac{n}{J}\theta_l^{short}\right) \\
& -2\frac{Jm\tau_n^2}{n^2}\sum_{l=1}^{n/m}\tau_n^2\left(\frac{n}{J}\widehat{\theta}_l^{short} - \frac{n}{J}\theta_l^{short}\right)\left(\frac{n}{m}\widehat{\theta}_l^{long} - \frac{n}{J}\theta_l^{short}\right).
\end{aligned}
$$

We have the following decomposition,

$$
\begin{aligned}
& \left(\frac{n}{J}\theta_l^{short} - \frac{n}{m}\theta_l^{long}\right)^2 \\
= \; & \left(\frac{n}{J}\int_{(l-1)m/n}^{[(l-1)m+J]/n} g(u)\,du - \frac{n}{m}\int_{(l-1)m/n}^{lm/n} g(u)\,du\right)^2 \\
\leq \; & \left(\frac{n}{J}\int_{(l-1)m/n}^{[(l-1)m+J]/n} \left(g(u) - g((l-1)m/n)\right)du\right)^2 + \left(\frac{n}{m}\int_{(l-1)m/n}^{lm/n} \left(g(u) - g((l-1)m/n)\right)du\right)^2 \\
& +2\left|\frac{n}{J}\int_{(l-1)m/n}^{[(l-1)m+J]/n} \left(g(u) - g((l-1)m/n)\right)du\right|\left|\frac{n}{m}\int_{(l-1)m/n}^{lm/n} \left(g(u) - g((l-1)m/n)\right)du\right|.
\end{aligned}
$$

These terms are small enough due to A4 and (B.1) as follows,

$$
\begin{aligned}
& \mathrm{E}\left|\frac{Jm\tau_n^2}{n^2}\sum_{l=1}^{K}\left(\frac{n}{m}\int_{(l-1)m/n}^{lm/n}\left(f(u) - f((l-1)m/n)\right)du\right)^2\right| & \text{(B.12)}\\
\leq \; & \frac{Jm\tau_n^2}{n^2}\sum_{l=1}^{K}\mathrm{E}\left(\frac{n}{m}\int_{(l-1)m/n}^{lm/n}\left(f(u) - f((l-1)m/n)\right)du\right)^2 \\
= \; & \frac{Jm\tau_n^2}{n^2}\sum_{l=1}^{K}\mathrm{E}\left(f(s_l) - f((l-1)m/n)\right)^2 \\
\leq \; & C\frac{Jm\tau_n^2}{n^2}\sum_{l=1}^{K}\mathrm{E}\left(\sigma(s_l) - \sigma((l-1)m/n)\right)^2 \\
\leq \; & C\frac{Jm\tau_n^2}{n^2}\sum_{l=1}^{K}\frac{m}{n} = C\frac{Jm\tau_n^2}{n^2} \to 0
\end{aligned}
$$

by assumption. In above, the first equality follows by mean value theorem, which applies by differentiability of

$$\int_{(l-1)m/n}^{t} \left(f(u) - f((l-1)\,m/n)\right) du \tag{B.13}$$

in time.

Next, we show

$$\frac{Jm\tau_n^2}{n^2}\sum_{l=1}^{K}\left(\frac{n}{m}\widehat{\theta}_l^{long} - \frac{n}{m}\theta_l^{long}\right)^2 \xrightarrow{p} 0.$$

By substituting m for J in

$$G^{(n)} = \frac{mJ}{n^2}\sum_{l=1}^{n/m}\tau_n^2\left(\frac{n}{J}\widehat{\theta}_l^{short} - \frac{n}{J}\theta_l^{short}\right)^2 \xrightarrow{p} V,$$

we obtain

$$\frac{m^2}{n^2}\sum_{l=1}^{n/m}\tau_n^2\left(\frac{n}{m}\widehat{\theta}_l^{long} - \frac{n}{m}\theta_l^{long}\right)^2 \xrightarrow{p} V,$$

and so by multiplying left hand side by J/m, (B.13) follows since $J/m \to 0$.

The remaining cross-terms in (B.10) are negligible by above results and Cauchy-Schwarz inequality. This concludes the proof of Theorem 2.4.1. ∎

B.2 Tables and Figures of Chapter 2

	m	800	2000	3000	2000	5000	7500	$\widehat{V}_a$
		$J = 200$			$J = 500$			
$\rho = 0$	*Two-sided*	0.96	0.97	0.98	0.92	0.95	0.95	0.93
	Left-sided	0.94	0.95	0.95	0.91	0.92	0.92	0.91
	Right-sided	0.97	0.98	0.98	0.96	0.97	0.97	0.96
$\rho = -0.3$	*Two-sided*	0.97	0.97	0.98	0.93	0.95	0.95	0.91
	Left-sided	0.96	0.97	0.97	0.93	0.95	0.95	0.91
	Right-sided	0.97	0.98	0.98	0.95	0.97	0.96	0.92
$\rho = -0.5$	*Two-sided*	0.97	0.98	0.98	0.93	0.95	0.96	0.84
	Left-sided	0.95	0.96	0.96	0.93	0.94	0.94	0.87
	Right-sided	0.97	0.98	0.98	0.95	0.97	0.97	0.91
$\rho = -0.7$	*Two-sided*	0.97	0.98	0.99	0.95	0.95	0.96	0.78
	Left-sided	0.95	0.96	0.96	0.92	0.93	0.94	0.81
	Right-sided	0.98	0.99	0.99	0.97	0.98	0.98	0.89

Table B.1: Coverage probabilities of 95% confidence interval of IV_X, $\xi^2 = 0.001$

	m	800	2000	3000	2000	5000	7500	$\widehat{V}_a$
		$J = 200$			$J = 500$			
$\rho = 0$	*Two-sided*	0.96	0.97	0.98	0.92	0.95	0.95	0.93
	Left-sided	0.94	0.95	0.95	0.91	0.92	0.92	0.91
	Right-sided	0.98	0.98	0.98	0.96	0.97	0.97	0.96
$\rho = -0.3$	*Two-sided*	0.97	0.98	0.98	0.93	0.95	0.95	0.93
	Left-sided	0.96	0.97	0.97	0.92	0.95	0.95	0.93
	Right-sided	0.97	0.98	0.98	0.95	0.97	0.96	0.95
$\rho = -0.5$	*Two-sided*	0.97	0.98	0.98	0.94	0.95	0.96	0.92
	Left-sided	0.95	0.96	0.96	0.93	0.94	0.94	0.91
	Right-sided	0.97	0.98	0.98	0.96	0.98	0.98	0.94
$\rho = -0.7$	*Two-sided*	0.97	0.98	0.98	0.94	0.96	0.96	0.88
	Left-sided	0.95	0.96	0.96	0.92	0.94	0.93	0.89
	Right-sided	0.98	0.99	0.99	0.96	0.98	0.98	0.94

Table B.2: Coverage probabilities of 95% confidence interval of IV_X, $\xi^2 = 0.0001$

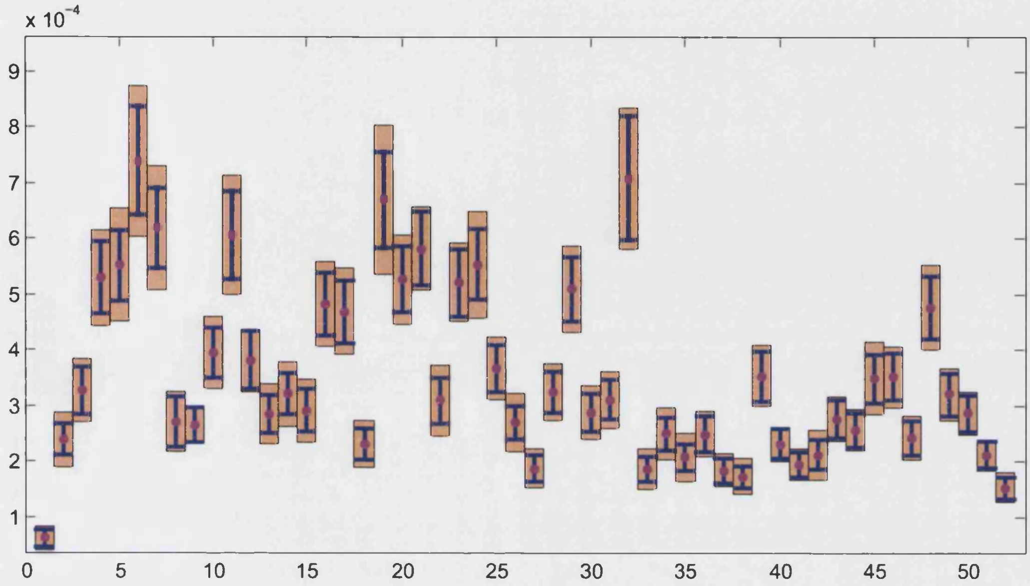


Figure B.1: *AIG stock*

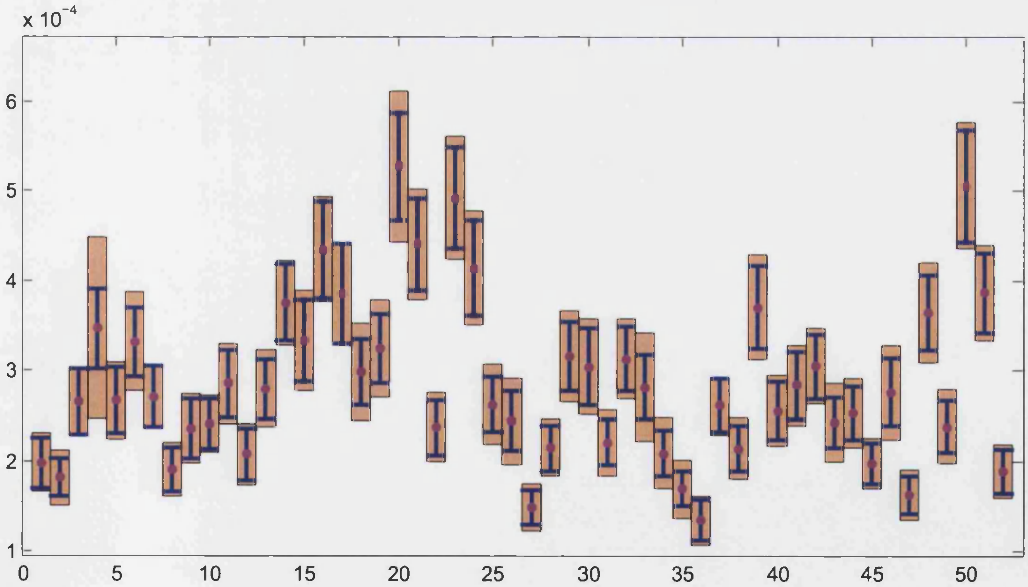


Figure B.2: *GE stock*

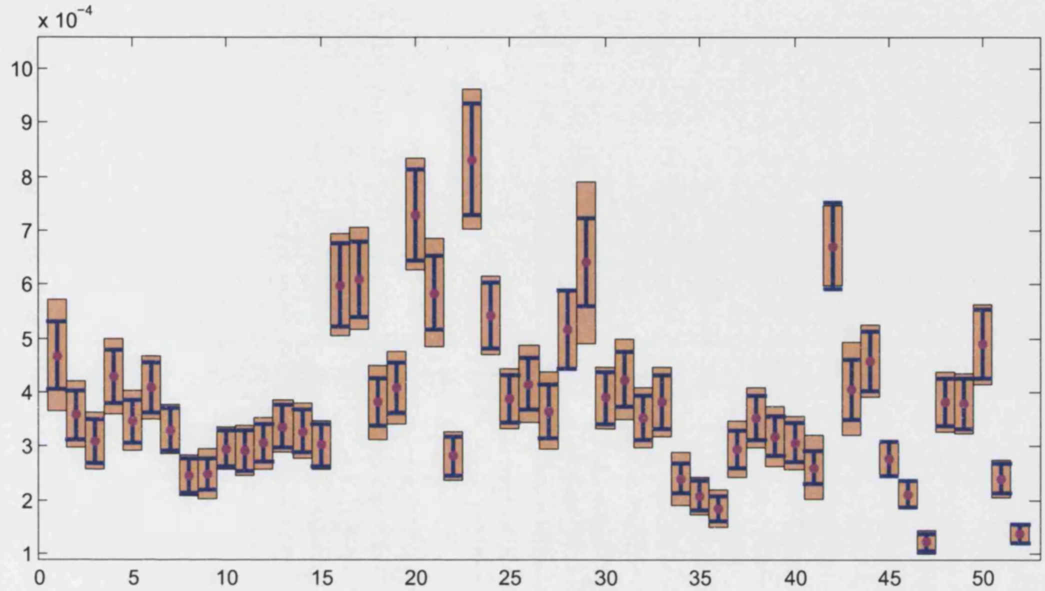


Figure B.3: *IBM stock*

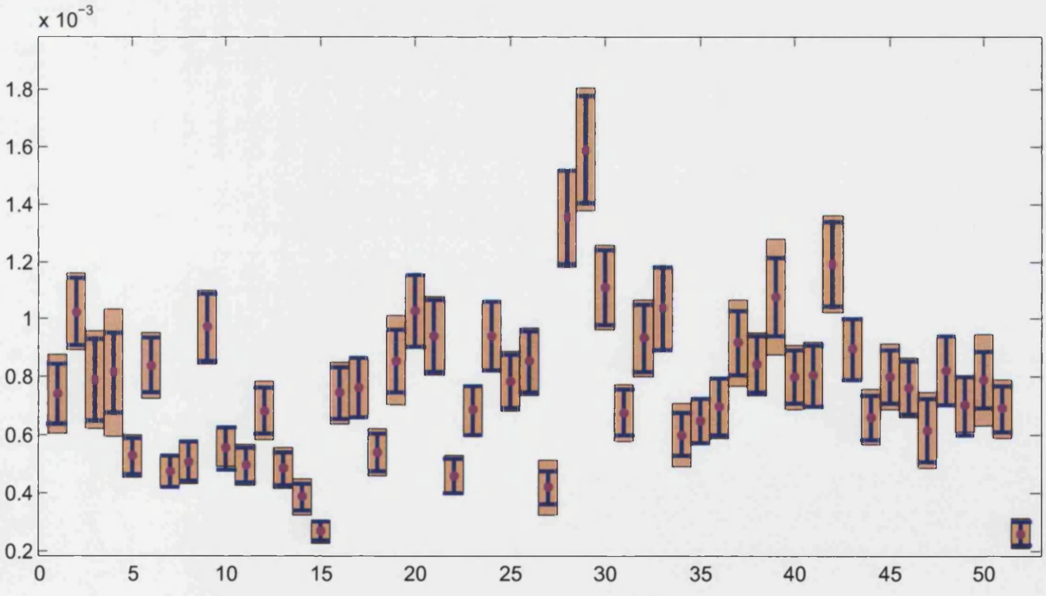


Figure B.4: *INTC stock*

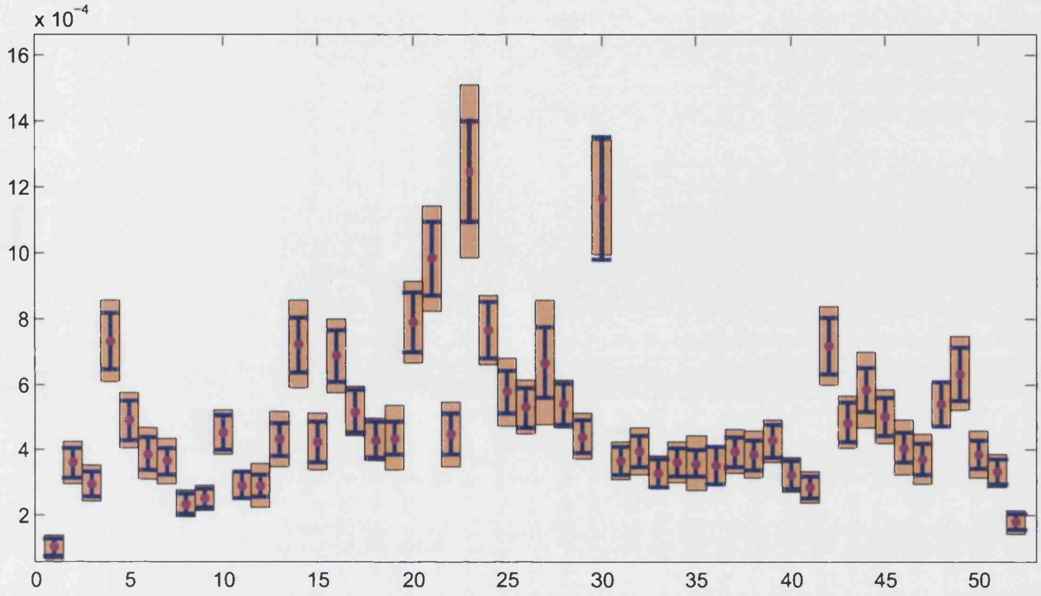


Figure B.5: *MMM stock*

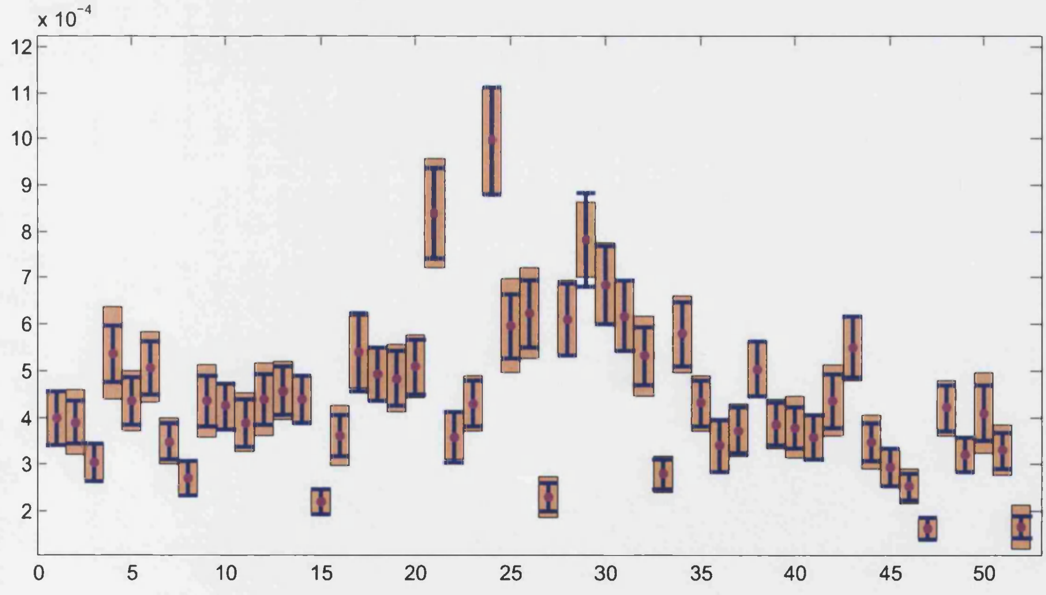


Figure B.6: *MSFT stock*

Appendix C

Tables and Figures of Chapter 3

C.1 Tables and Figures of Chapter 3

	trans./week	$\widehat{\omega}^2 \cdot 10^7$	$\widetilde{\omega}^2 \cdot 10^7$	$\widehat{\xi}^2 \cdot 10^5$	acf(1)	acf(2)	acf(3)
AIG	18,029	0.207	0.136	0.156	-0.320	0.102	-0.014
GE	29,015	0.228	0.188	0.189	-0.582	0.248	-0.118
IBM	20,070	0.162	0.095	0.117	-0.302	0.081	0.008
INTC	35,267	0.518	0.407	0.127	-0.525	0.200	-0.085
MMM	14,005	0.284	0.123	0.121	-0.269	0.092	0.006
MSFT	32,421	0.338	0.282	0.178	-0.555	0.224	-0.100
SPY	39,801	0.037	0.018	0.048	-0.352	0.065	0.006

Table C.1: *Summary statistics of data before synchronization. First column contains average number of transactions per week. Second and third columns contains variance of the noise estimates over the whole year 2006,* $\widehat{\omega}^2 = RV/2n$, $\widehat{\omega}^2 = \left(RV - \widehat{IV}\right)/2n$ *where IV is estimated by the TSRV; n is total number of transactions in 2006 for the corresponding stock. Fourth column contains estimated noise-to-signal ratio,* $\widehat{\xi}^2 = \widehat{\omega}^2/\widehat{IV}$. *Last three columns contain autocorrelation functions of returns at first, second, and third lag.*

	trans./week	$\widehat{\omega}^2 \cdot 10^7$	$\widetilde{\omega}^2 \cdot 10^7$	$\widehat{\xi}^2 \cdot 10^5$	acf(1)	acf(2)	acf(3)
AIG(SPY)	15,425	0.220	0.138	0.282	-0.15	0.051	0.02
GE(SPY)	21,819	0.229	0.176	0.295	-0.221	0.058	0.015
IBM(SPY)	16,890	0.174	0.095	0.223	-0.166	0.052	0.021
INTC(SPY)	24,601	0.545	0.384	0.700	-0.247	0.060	0.016
MMM(SPY)	12,315	0.303	0.121	0.389	-0.114	0.048	0.014
MSFT(SPY)	23,322	0.347	0.267	0.451	-0.238	0.061	0.017
SPY(AIG)	15,425	0.059	0.011	0.045	-0.276	0.084	-0.006
SPY(GE)	21,819	0.049	0.014	0.040	-0.509	0.173	-0.059
SPY(IBM)	16,890	0.056	0.011	0.040	-0.257	0.069	0.011
SPY(INTC)	24,601	0.045	0.014	0.011	-0.439	0.132	-0.041
SPY(MMM)	12,315	0.071	0.011	0.031	-0.232	0.082	0.013
SPY(MSFT)	23,322	0.046	0.014	0.024	-0.476	0.155	-0.051

Table C.2: *Summary statistics of data after synchronization. Notation AIG(SPY) means stock AIG after it has been synchronized with SPY. By construction, number of transactions of AIG(SPY) is the same as that of SPY(AIG). See Table 1 annotation for meaning of other column entries.*

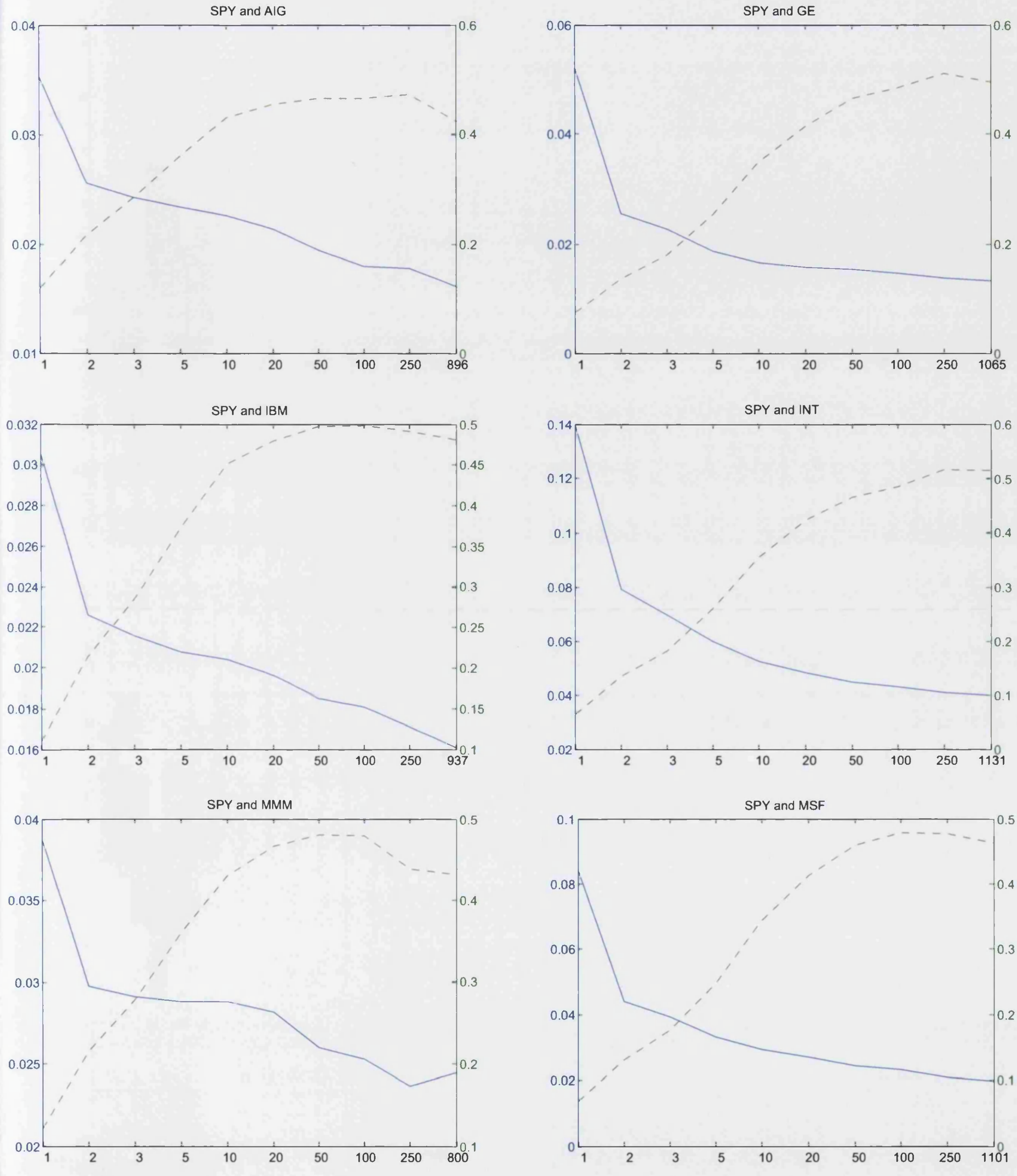


Figure C.1: *The solid lines (left y-axis) are the volatility signature plots, i.e., realized variance plotted against the frequency (in ticks) used in its calculation. Dashed lines (right y-axis) are the realized covariance plots against the frequency (in ticks). Data covers the whole year 2006.*

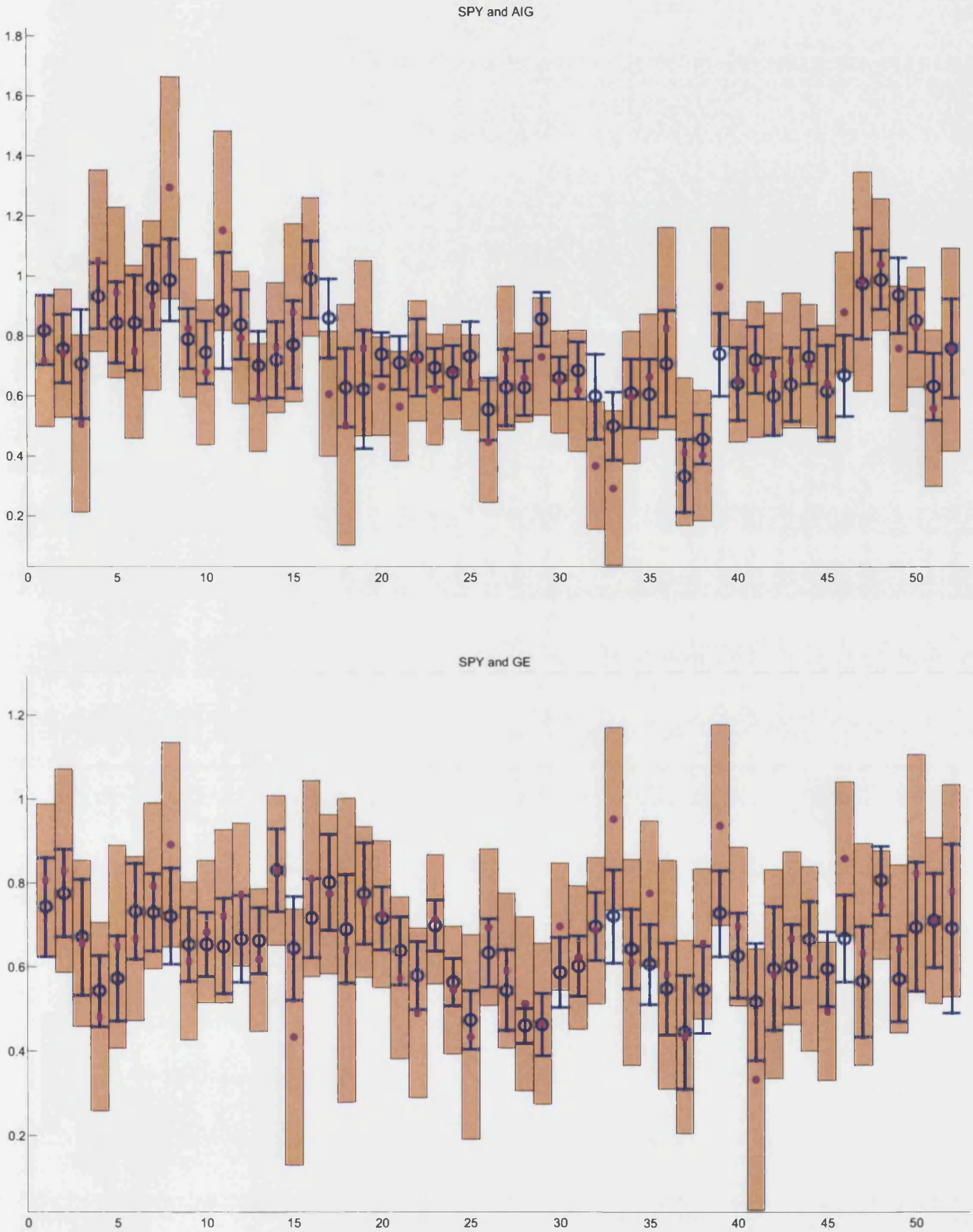


Figure C.2: *Estimated betas for AIG and GE with 95% confidence intervals. Filled dots with rectangular CIs correspond to* $\widehat{\beta}_{5min}^{RV}$, *empty dots with error-bar-type CIs correspond to* $\widehat{\beta}^{AMZ}$. *Weeks on the x-axis.*

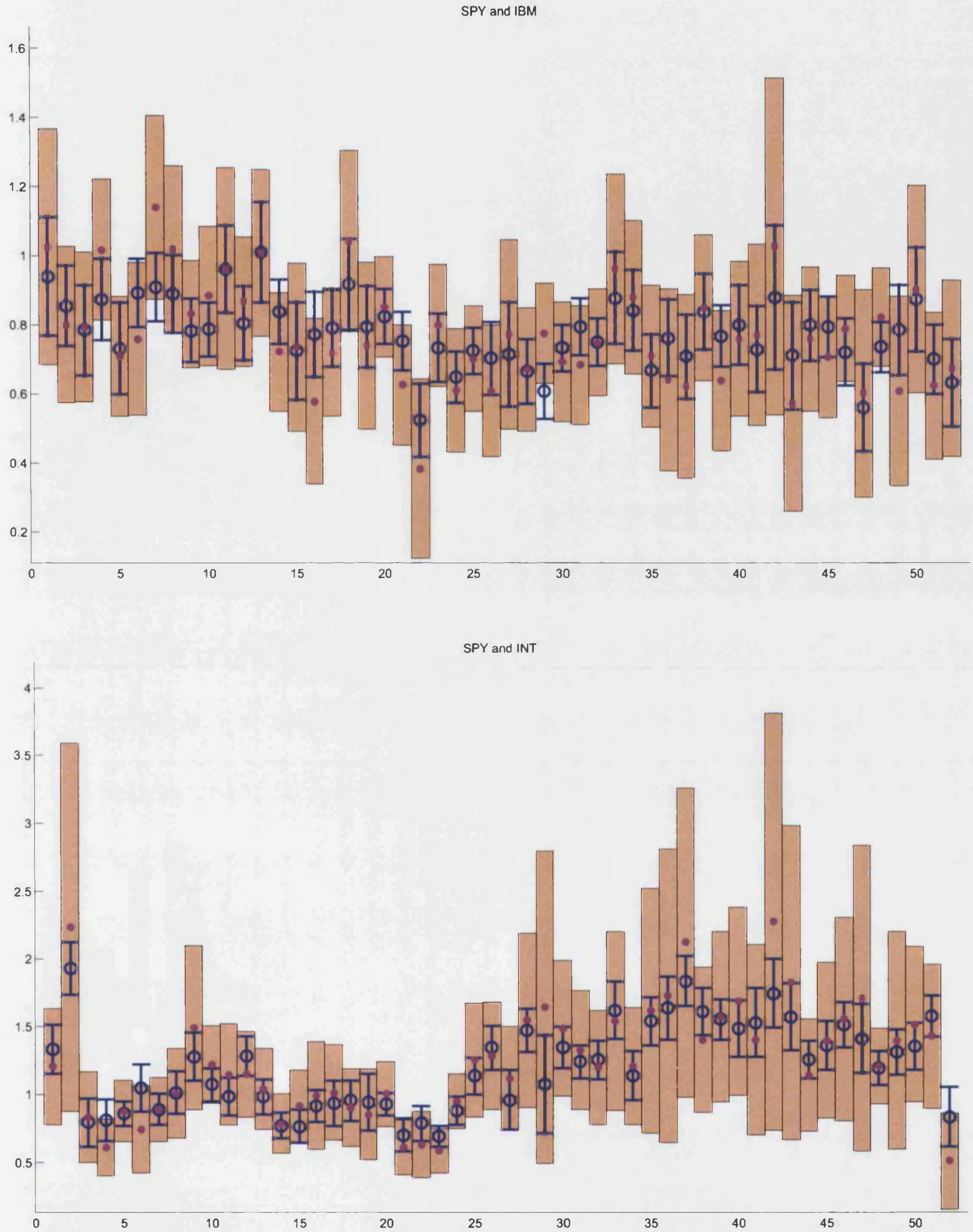


Figure C.3: *Estimated betas for IBM and INTC with 95% confidence intervals. Filled dots with rectangular CIs correspond to* $\widehat{\beta}_{5min}^{RV}$, *empty dots with error-bar-type CIs correspond to* $\widehat{\beta}^{AMZ}$. *Weeks on the x-axis.*

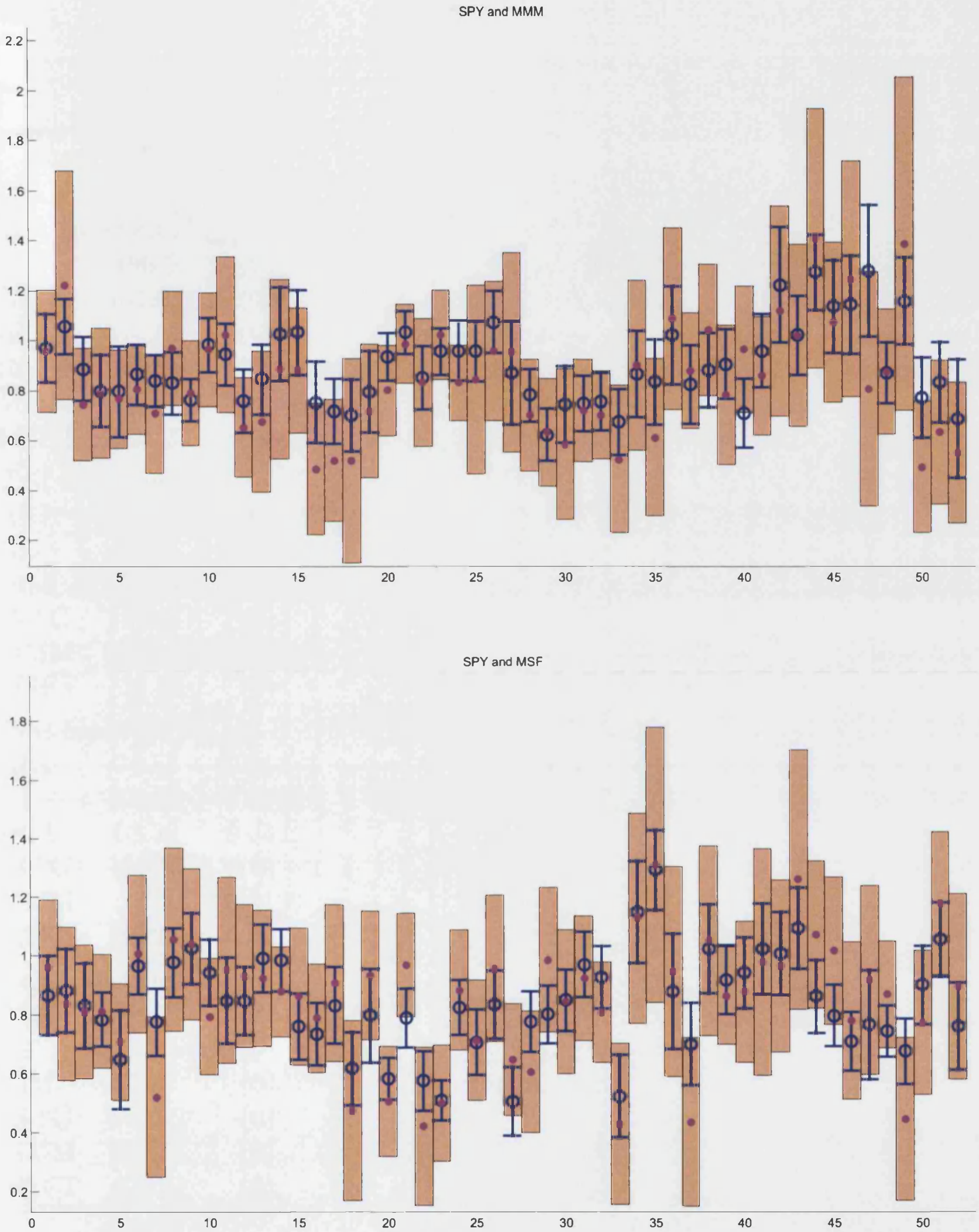


Figure C.4: *Estimated betas for MMM and MSFT with 95% confidence intervals. Filled dots with rectangular CIs correspond to* $\widehat{\beta}^{RV}_{5min}$ *, empty dots with error-bar-type CIs correspond to* $\widehat{\beta}^{AMZ}$ *. Weeks on the x-axis.*

	2006		Q1		Q2		Q3		Q4	
Test based on $\widehat{\beta}^{RV}_{5min}$										
AIG	196.5	(0)	23.47	(0.024)	35.63	(0)	52.83	(0)	56.28	(0)
GE	101.03	(0)	15.26	(0.228)	15.55	(0.213)	27.75	(0.006)	26	(0.011)
IBM	110.82	(0)	14.58	(0.265)	30.02	(0.003)	26.59	(0.009)	17.41	(0.135)
INTC	248.96	(0)	33.84	(0.001)	40.96	(0)	36.15	(0)	43.65	(0)
MMM	132.45	(0)	27.48	(0.007)	25.1	(0.014)	21.83	(0.039)	37.48	(0)
MSFT	213.3	(0)	22.74	(0.03)	71.43	(0)	64.97	(0)	42.23	(0)
Test based on $\widehat{\beta}^{RV}_{15min}$										
AIG	112.7	(0)	26.03	(0.011)	19.99	(0.067)	33.69	(0.001)	15.78	(0.202)
GE	75.14	(0.012)	12.23	(0.427)	19.81	(0.071)	25.06	(0.015)	17.17	(0.143)
IBM	73.48	(0.017)	15.09	(0.236)	21.33	(0.046)	8.94	(0.708)	8.57	(0.739)
INTC	134.91	(0)	27.62	(0.006)	30.37	(0.002)	6.62	(0.882)	24.29	(0.019)
MMM	85.34	(0.001)	14.52	(0.269)	25.03	(0.015)	16.88	(0.154)	28.24	(0.005)
MSFT	121.52	(0)	15.49	(0.216)	39.48	(0)	34.18	(0.001)	22.8	(0.029)
Test based on $\widehat{\beta}^{RV}_{20min}$										
AIG	74.54	(0.014)	24.37	(0.018)	9.43	(0.666)	9.19	(0.687)	22.8	(0.03)
GE	85.89	(0.001)	16.26	(0.18)	15.16	(0.233)	17.38	(0.136)	22.89	(0.029)
IBM	68.38	(0.043)	10.93	(0.535)	17.8	(0.122)	11.85	(0.458)	7.36	(0.833)
INTC	107.53	(0)	20.81	(0.053)	29.48	(0.003)	3.36	(0.992)	19.16	(0.085)
MMM	73	(0.019)	14.54	(0.268)	14.3	(0.282)	16.99	(0.15)	20.9	(0.052)
MSFT	87.09	(0.001)	13.61	(0.326)	19.62	(0.075)	28.56	(0.005)	15.33	(0.224)
Test based on $\widehat{\beta}^{AMZ}$										
AIG	269.04	(0)	22.44	(0.033)	34.97	(0)	71.28	(0)	57.31	(0)
GE	224.4	(0)	20.56	(0.057)	65.13	(0)	63.42	(0)	26.08	(0.01)
IBM	136.7	(0)	18.93	(0.09)	37.52	(0)	26.76	(0.008)	18.96	(0.09)
INTC	845.32	(0)	143.45	(0)	83.1	(0)	79.34	(0)	55.97	(0)
MMM	200.38	(0)	28.75	(0.004)	40.19	(0)	21.96	(0.038)	63.09	(0)
MSFT	403.21	(0)	30.24	(0.003)	96.47	(0)	124.55	(0)	60.15	(0)

Table C.3: *Values of the Chi-square test; corresponding p-values in parenthesis. The null hypothesis is that true betas are constant over the some time interval. The top row indicates the corresponding time interval.*