Quasi-Maximum Likelihood Estimation of Volatility with High Frequency Data*

Dacheng Xiu

First draft: Sep 2008. This version: May 2010

Abstract

This paper investigates the properties of the well-known maximum likelihood estimator in the presence of stochastic volatility and market microstructure noise, by extending the classic asymptotic results of quasi-maximum likelihood estimation. When trying to estimate the integrated volatility and the variance of noise, this parametric approach remains consistent, efficient and robust as a quasi-estimator under misspecified assumptions. Moreover, it shares the model-free feature with nonparametric alternatives, for instance realized kernels, while being advantageous over them in terms of finite sample performance. Comparisons with a variety of implementations of the Tukey-Hanning$_2$ kernel are provided using Monte Carlo simulations, and an empirical study with the Euro/US Dollar future illustrates its application in practice.

Key Words: Integrated volatility, Market microstructure noise, Quasi-Maximum Likelihood Estimator, Realized Kernels, Stochastic volatility.

JEL classification: C13; C22; C51

*I am indebted to Yacine Aït-Sahalia for encouraging me to work on this topic and helpful suggestions thereafter. I appreciate the comments of the editor and two referees, which led to considerable improvements of the paper. I am also grateful to Jianqing Fan, Peter Reinhard Hansen, Jean Jacod, Yingying Li, Ulrich K. Müller, Per Mykland, Eric Renault, and Ronnie Sircar for valuable comments, which helped a lot in improving the quality of the paper. Also, I would like to thank Todd Hines for data assistance, Jia Li and Xin Wan as well as the participants of the seminars at Princeton University and the University of Chicago, of the workshop on Financial Econometrics at Fields Institute, and of the 2nd Princeton-Humboldt Finance workshop for fruitful discussions. All errors are mine.

1 Introduction

The availability of high frequency data is a double-edged sword for the estimation of volatility. On the one hand, it facilitates our empirical studies of the asymptotic properties of a natural estimator, Realized Variance (RV), i.e. the sum of squared log-returns; while on the other hand, along with the data comes market microstructure noise, which disrupts all the desirable properties of the estimator. Without microstructure noise, this simple estimator is both consistent and efficient. In addition, other studies have shed light on its central limit theory, such as Jacod (1994) and Barndorff-Nielsen and Shephard (2002). However, the existence of noise interferes with the estimation especially when the sampling frequency approaches zero. A common practice is to sample sparsely, say every 5 minutes, discarding a large portion of the sample and the information therein.

The problem has received considerable attention recently. For instance, Aït-Sahalia et al. (2005) suggest sampling as frequently as possible at the cost of modeling the noise. This paper assumes constant volatility so that it can perform the maximum likelihood estimation (MLE). In the setting of stochastic volatility, Zhang et al. (2005) bring forward a nonparametric estimator, Two-Scale Realized Volatility (TSRV), which is the first consistent estimator in the presence of noise, despite a relatively low convergence rate $n^{\frac{1}{6}}$. Subsequently, Zhang (2006) advocates Multi-Scale Realized Volatility (MSRV), improving the convergence rate to $n^{\frac{1}{4}}$, which is the optimal rate a model can reach as shown by Gloter and Jacod (2001). Recently, Barndorff-Nielsen et al. (2008) have designed various Realized Kernels (RKs) that can be used to deal with endogenous noise and endogenously spaced data and their convergence rates are the same as that of the MSRV. In simulation, these nonparametric estimators perform very well with optimally selected bandwidth or kernels; while in practice, estimators with bandwidth based on ad hoc choices, or based on small sample performance may behave better, as illustrated in Gatheral and Oomen (2007) and Bandi and Russell (2008). Another group of estimators dates back to Zhou (1996), who first applies the autocovariances-based-approach to constant volatility cases. Hansen and Lunde (2006) extend the Zhou estimator to the stochastic volatility models with serially dependent noise. However, the Zhou estimator and its extensions are inconsistent.

The motivation and inspiration of this article stem from Gatheral and Oomen (2007) and Aït-Sahalia and Yu (2009). Using artificially simulated "zero-intelligence" data, Gatheral and

Oomen (2007) compare a comprehensive set of estimators including those listed above and their ad hoc modifications. According to their studies, the MLE, though possibly misspecified with time varying volatility, is among the best in terms of efficiency and robustness. In addition, Aït-Sahalia and Yu (2009) apply the MLE to analyze the liquidity of NYSE stocks. Their maximum likelihood estimates with the data simulated from stochastic volatility models indicate good properties, although this estimator is derived from a constant volatility assumption. Related works also include Hansen et al. (2008), where the authors suggest the consistency of the MLE by examining moving average filters. These studies motivate us to consider the MLE as a Quasi-Maximum Likelihood Estimator (QMLE)[1] under misspecified models, which dates back to as early as Amemiya (1973), White (1980) and White (1982). In these articles, the framework of misspecified estimation has been built and its close connection with Kullback-Leibler Information Criterion (KLIC) (see Kullbacks and Leibler (1951)) has been illustrated. Since these seminal works, Domowitz and White (1982) and Bates and White (1985) have extended the consistency results of the QMLE with i.i.d. models to various cases including dependent observations, Quasi-GMM-estimators, and Quasi-M-estimators.

Our work is thereby built on the fusion of high frequency data and misspecified likelihood estimation. The correct model specification features stochastic volatility. However, this model is intentionally misspecified to be one of constant volatility. Under this assumption, we perform (quasi) maximum likelihood estimation and analyze the estimator, which is essentially the same as the MLE in Aït-Sahalia et al. (2005). Remarkably, in the context of the correct model, the QMLE of volatility consistently estimates the integrated volatility at the most efficient rate $n^{\frac{1}{4}}$. Also, the QMLE of noise variance has the same asymptotic distribution as before. In other words, the maximum likelihood estimators are robust to stochastic volatility. In addition, they are still robust to random sampling intervals and non-Gaussian market microstructure noises.

The paper is organized as follows. Section 2 reviews the parametric likelihood estimator and provides the intuition and motivation for the QMLE. Section 3 outlines the classic asymptotic theory of the QMLE, and derives an extension to more general settings. Section 4 investigates the statistical properties of the QMLE, where the consistency, central limit theory and robustness of the QMLE are established. Section 5 compares it with nonparametric ker-

[1] We give the MLE an alias QMLE sometimes in order to emphasize model misspecification and keep the notation in line with the classic results of misspecified models.

nel estimators, and Section 6 uses Monte Carlo simulations to verify the conclusions obtained from the previous sections. Section 7 details an empirical study with the Euro/US Dollar future data. Section 8 concludes. The appendix provides all mathematical proofs.

2 Revisiting the MLE: the QMLE

In this section, we recapitulate the parametric approach and introduce our quasi-estimator. The traditional parametric methodology applies to cases where the true value of the parameter of interest is a special point in the parameter space. Therefore, Aït-Sahalia et al. (2005) have to make an assumption that the volatility is neither a stochastic process nor a deterministic function, but instead, a constant. That is, the latent efficient log price process satisfies

$$dX_t = \sigma dW_t \tag{1}$$

with the observed log transaction price $\tilde{X}_{\tau_i}$ contaminated by the microstructure noise U in a way such that $\tilde{X}_{\tau_i} = X_{\tau_i} + U_{\tau_i}$, where τ_i is the observation time. For simplicity, we assume that the data are regularly spaced, satisfying $\tau_i - \tau_{i-1} = \Delta$. The observations are made within $[0, T]$, where $T = n \cdot \Delta$ is fixed, so that the infill asymptotic behaviors are determined as n goes to ∞ and Δ goes to 0 simultaneously. The structure of the observed log return Y_is features MA(1), where

$$Y_i = \tilde{X}_{\tau_i} - \tilde{X}_{\tau_{i-1}} = \sigma(W_{\tau_i} - W_{\tau_{i-1}}) + U_{\tau_i} - U_{\tau_{i-1}}$$

If we postulate that the noise distribution is Gaussian, then our log likelihood function for Ys is

$$l(\sigma^2, a^2) = -\frac{1}{2}\log\det(\Omega) - \frac{n}{2}\log(2\pi) - \frac{1}{2}Y'\Omega^{-1}Y \tag{2}$$

where

$$\Omega = \begin{pmatrix} \sigma^2\Delta + 2a^2 & -a^2 & 0 & \cdots & 0 \\ -a^2 & \sigma^2\Delta + 2a^2 & -a^2 & \ddots & \vdots \\ 0 & -a^2 & \sigma^2\Delta + 2a^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -a^2 \\ 0 & \cdots & 0 & -a^2 & \sigma^2\Delta + 2a^2 \end{pmatrix} \tag{3}$$

The MLE $(\hat{\sigma}^2, \hat{a}^2)$ proves to be consistent at different rates for its volatility part and noise part even if the noise distribution turns out to be non-Gaussian:

$$\begin{pmatrix} n^{\frac{1}{4}}(\hat{\sigma}^2 - \sigma_0^2) \\ n^{\frac{1}{2}}(\hat{a}^2 - a_0^2) \end{pmatrix} \xrightarrow{\mathcal{L}} N\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 8a_0\sigma_0^3 T^{-\frac{1}{2}} & 0 \\ 0 & 2a_0^4 + \mathrm{cum}_4[U] \end{pmatrix} \Big)$$

where $\mathrm{cum}_4[U]$ is the fourth cumulant of the true noise U.

Nevertheless, ample evidence of stochastic volatility calls this simplistic model into question. So, it is natural to ask: what is the impact of stochastic volatility on the MLE? Will stochastic volatility change the desired properties of the MLE, just as microstructure noise does to RV, or will the MLE still be consistent? Simulation studies by Aït-Sahalia and Yu (2009) and Hansen et al. (2008) seem to have suggested that the MLE may be a consistent estimator of the integrated volatility. Intuitively, this conjecture is plausible in that when volatility becomes stochastic, the integrated volatility, the parameter of interest, happens to be the average of the volatility process, which is expected to be a legitimate candidate for the estimate. If consistency is guaranteed, what would be the convergence rate and asymptotic variance? How would it compare with other alternative nonparametric estimators? Closer scrutiny of the estimator in the absence of microstructure noise may yield more insights.

Consider a stochastic volatility model with no noise:

$$dX_t = \sigma_t dW_t$$

The objective is to estimate the integrated volatility $\int_0^T \sigma_t^2 dt$. By design, we mistakenly assume the spot volatility σ_t to be constant; therefore, the quasi-log likelihood function is

$$l(\omega, \sigma^2) = -\frac{n}{2}\log(\sigma^2\Delta) - \frac{n}{2}\log(2\pi) - \frac{1}{2\sigma^2\Delta}Y'Y$$

where $Y_i = X_{\tau_i} - X_{\tau_{i-1}} = \int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t$ and $Y = (Y_1, Y_2, \ldots, Y_n)'$.

Apparently, the QMLE is

$$\hat{\sigma}_n^2 = \frac{1}{T}\sum_{i=1}^{n} Y_i^2 = \frac{1}{T}\sum_{i=1}^{n}(X_{\tau_i} - X_{\tau_{i-1}})^2$$

Here, the RV estimator recurs through a different argument, and it is of course the perfect estimator under the true stochastic volatility model.

However, the consistency of the QMLE is no longer straightforward in the presence of noise, because there may be no close form available for this estimator. Its asymptotic variance is far more complicated due to heteroskedasticity and autocorrelation, as mentioned by Hansen et al. (2008). The remaining paper rigourously investigates the asymptotic behavior of the QMLE in the setting of stochastic volatility and microstructure noise, which leads us to the classic asymptotic theory of quasi-estimators.

3 Asymptotic Theory of Quasi-M-Estimators

In this section, we at first briefly review the consistency of the QMLE under misspecified models, a theory initially developed in White (1980) and White (1982). The rationale behind the theory is related to Kullback - Leibler Information Criterion (KLIC). More precisely, suppose that we have an i.i.d. random sample, and let $g(X)$ be the true unknown data generating density, and $f(X, \theta)$ our possibly misspecified density indexed by a parameter $\theta \in \Theta$. White (1982) claims that under certain regular conditions, the QMLE is consistent to θ^* which minimizes KLIC:

$$I(g : f, \theta) = E(\log[g(X)/f(X, \theta)])$$

where the expectation is taken under the true model. If the model is correctly specified, that is, there exists $\theta_0 \in \Theta$, such that $g(X) = f(X, \theta_0)$, then the KLIC attains its minimum at $\theta^* = \theta_0$, hence this result is in agreement with the consistency of regular maximum likelihood estimators. Otherwise, the model is misspecified, and intuitively, θ^* minimizes our ignorance of the true structure.

In Domowitz and White (1982), the authors generalize these results to include dependent observations, and show that the QMLE $\hat{\theta}_n$, which maximizes $Q_n(\omega, \theta)$, still converges in probability to θ_n^*[2], a maximizer of $\bar{Q}_n(\omega, \theta)$, the expectation of $Q_n(\omega, \theta)$ under the true model. In addition, Bates and White (1985) extend the theory to general quasi-M-estimators, by introducing discrepancy functions.

Now, we add more randomness to the misspecification theory, which enables our applications to stochastic volatility models, where the parameter of interest itself is random. The reasoning of the proof is similar to the regular one given in White (1980) and Newey and McFadden (1994).

[2] θ^* might depend on n when model is misspecified.

Theorem 1. *Let $Q_n(\omega,\theta)$ and $\bar{Q}_n(\omega,\theta)$ be two random functions such that for each θ in Θ, a compact subset of R^k, they are measurable functions on Ω and, for each $\omega \in \Omega$, continuous functions on Θ. In addition, $\bar{Q}_n(\omega,\theta)$ is almost surely maximized at $\theta_n^*(\omega)$. Further, the following two conditions are satisfied as $n \to \infty$:*

1. *Uniform Convergence:*
$$\sup_{\theta\in\Theta} \|Q_n(\omega,\theta) - \bar{Q}_n(\omega,\theta)\| \xrightarrow{\text{P}} 0. \tag{4}$$
2. *Identifiability: for every $\epsilon > 0$, there exists a constant $\delta_0 > 0$, such that*
$$P(\bar{Q}_n(\omega,\theta_n^*) - \max_{\theta\in\Theta:\|\theta-\theta_n^*\|\geq\epsilon} \bar{Q}_n(\omega,\theta) > \delta_0) \to 1. \tag{5}$$

Then any sequence of estimators $\hat{\theta}_n$ such that $Q_n(\omega,\hat{\theta}_n) \geq \sup_{\theta\in\Theta} Q_n(\omega,\theta) + o_p(1)$ converges in probability to θ_n^*, i.e., $\hat{\theta}_n - \theta_n^* \xrightarrow{\text{P}} 0$.

Note that the identifiability condition here is slightly different from that in a common setting, for example Van Der Vaart (2000) and Newey and McFadden (1994). It not only requires some uniqueness like property of the maximizer, but also a proper normalization such that the maximizer can be distinguished asymptotically.

Now we modify the assumptions to accommodate to the M-Estimators setting, in case we need different normalizations for different parameters.

Theorem 2. *Let $\Psi_n(\omega,\theta)$and $\bar{\Psi}_n(\omega,\theta)$ be random vector-valued functions. For each θ in Θ, a compact subset of $\mathbb{R}^k$, they are measurable function on Ω, and for each ω in Ω, continuous functions on Θ. In addition, there exists a sequence of θ_n^*, satisfying $\bar{\Psi}_n(\omega,\theta_n^*) = 0$ almost surely, such that as $n \to \infty$,*

1. *Uniform Convergence:*
$$\sup_{\theta\in\Theta} \|\Psi_n(\omega,\theta) - \bar{\Psi}_n(\omega,\theta)\| \xrightarrow{\text{P}} 0. \tag{6}$$
2. *Identifiability: For every $\epsilon > 0$, there exists a constant $\delta_0 > 0$, such that,*
$$P(\min_{\theta\in\Theta:\|\theta-\theta_n^*\|\geq\epsilon} \|\bar{\Psi}_n(\omega,\theta)\| > \delta_0) \to 1. \tag{7}$$

Then any sequence of estimators $\hat{\theta}_n$ such that $\Psi_n(\omega,\hat{\theta}_n) = o_p(1)$ converges in probability to θ_n^*, i.e., $\hat{\theta}_n - \theta_n^* \xrightarrow{\text{P}} 0$.

In the following discussions, we will choose Ψ_n as the score function of a misspecified model, up to an appropriate normalization, and $\bar{\Psi}_n$ is carefully specified corresponding to Ψ_n.

The central limit result is given by the next theorem, which is an extension of Theorem 2.4 in Domowitz and White (1982).

Theorem 3. *Suppose that the conditions of Theorem 2 are satisfied. In addition, $\Psi_n(\omega,\theta)$ and $\bar{\Psi}_n(\omega,\theta)$ are continuously differentiable of order 1 on Θ. Also, there exists a sequence of positive definite matrices $\{V_n(\omega)\}$ such that*

$$-V_n(\omega)\Psi_n(\omega,\theta_n^*) \xrightarrow{\mathcal{L}} N(0, I_k) \tag{8}$$

If $\nabla\bar{\Psi}_n(\omega,\theta)$ is stochastic equicontinuous, and $|\nabla\Psi_n(\omega,\theta) - \nabla\bar{\Psi}_n(\omega,\theta)| \xrightarrow{\mathrm{P}} 0$, uniformly for all $\theta \in \Theta$, then

$$V_n(\omega)\nabla\bar{\Psi}_n(\omega,\theta_n^*)(\hat{\theta}_n - \theta_n^*) \xrightarrow{\mathcal{L}} N(0, I_k) \tag{9}$$

The extensions of the classic asymptotic theory pave the way for a thorough inquiry of asymptotic properties of the QMLE.

4 Statistical Properties of the QMLE

This section shares the setup with most volatility estimators available in the literature. More specifically, we make the following assumptions.

Assumption 1. *The underlying latent log price process satisfies*

$$dX_t = \sigma_t dW_t$$

with the volatility process a positive and locally bounded Itô semimartingale.[3]

Assumption 2. *The noise U_t is independently and identically distributed, and independent of price and volatility processes, with mean 0, variance a_0^2 and finite fourth moment.*

The assumptions of i.i.d. noise and its independence with prices are not always consistent with the empirical evidence, as pointed out in Hansen and Lunde (2006). Aït-Sahalia et al.

[3]This assumption accommodates virtually all continuous time financial models. See Jacod (2008) Hypothesis (L-s) for more details.

(2005) have discussed the way to modify the log likelihood function with more parameters, according to the assumed parametric structure of the noise. See also Gatheral and Oomen (2007) for an MA(2) implementation. As to the TSRV, Aït-Sahalia et al. (2006) have extended it to the case with serially dependent noise. Kalnina and Linton (2008) have proposed a modification of the TSRV with endogenous noise and heteroskedastic measurement error. Independently, Barndorff-Nielsen et al. (2008) have shown the robustness of their realized kernels with respect to endogenous noise. However, serially dependence plus endogenous assumption are still unsatisfactory, since in reality the transaction prices are recorded with round-off errors. Li and Mykland (2007) have discussed the robustness of the TSRV with respect to rounding errors while Jacod et al. (2007) have recently proposed a pre-averaging approach that works well for a general class of errors including certain combination of rounding and additive errors. In this paper, being parsimonious and for simplicity, we consider the i.i.d. white noise and provide in addition a heuristic argument for time-dependent noise to validate the applications of the QMLE in practice.

4.1 Consistency of the QMLE

Consider first what would happen in the absence of noise. Based on the discussion in Section 2, the score function (up to a proper normalization) under misspecified model is,

$$\Psi_n(\omega, \sigma^2) = -\frac{1}{n}\frac{dl(\omega, \sigma^2)}{d\sigma^2} = \frac{1}{2}\{\frac{1}{\sigma^2} - \frac{1}{n\sigma^4\Delta}Y'Y\}$$

and its root is

$$\hat{\sigma}_n^2 = \frac{1}{T}\sum_{i=1}^{n} Y_i^2 = \frac{1}{T}\sum_{i=1}^{n}(X_i - X_{i-1})^2$$

Then we choose

$$\bar{\Psi}_n(\omega, \sigma^2) = \frac{1}{2}\{\frac{1}{\sigma^2} - \frac{1}{n\sigma^4\Delta}\int_0^T \sigma_t^2 dt\} \tag{10}$$

Therefore, it has a root $\sigma_n^{2*} = \frac{1}{T}\int_0^T \sigma_t^2 dt$.

Because σ^2 is in a compact set Θ, and if we require the parameter space to be bounded away from zero, i.e., there exist $\tilde{\sigma}^2 \in \Theta$ such that $\sigma^2 \geq \tilde{\sigma}^2 > 0$, then

$$\sup_{\sigma^2\in\Theta} |\Psi_n(\omega, \sigma^2) - \bar{\Psi}_n(\omega, \sigma^2)| = \frac{1}{2\tilde{\sigma}^4 T}|\sum_{i=1}^{n} Y_i^2 - \int_0^T \sigma_t^2 dt| \xrightarrow{\text{P}} 0 \tag{11}$$

Proof of the last step is well-known, see e.g. Karatzas and Shreve (1991, pp32 Theorem 5.8), hence uniform convergence in probability in Theorem 2 is shown. Identifiability condition trivially holds, so the consistency follows from Theorem 2. The rate of convergence depends on the rate of (11), which is $n^{\frac{1}{2}}$, as given by Jacod (1994) for instance. Therefore,

$$\hat{\sigma}_n^2 - \sigma_n^{2*} = \frac{1}{T}\sum_{i=1}^n Y_i^2 - \frac{1}{T}\int_0^T \sigma_t^2 dt = O_p(n^{-\frac{1}{2}})$$

Apparently, we do not need Theorem 2 at all here. In general, however, this is certainly not the case. One inspiration from this simple example is regarding the selection of $\bar{\Psi}_n$ in (10). If we add no leverage assumption, that is to say, the volatility process is conditionally deterministic, then $\bar{\Psi}_n$ is nothing but the conditional expectation of Ψ_n under the true model.

When the observed data are noisy, we have $Y_i = \int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t + U_{\tau_i} - U_{\tau_{i-1}}$. Besides the constant volatility assumption, we mistakenly assume that the noises are normally distributed with variance a^2. Therefore, the quasi-log likelihood function is exactly (2), hence the same likelihood estimator recurs in the form of the QMLE under misspecified model. However, closed-form expressions are no longer available, so we have to turn to Theorem 2 for consistency.

Denote $\theta = (\sigma^2, a^2)$ and we also assume that the parameters stay in a compact set, which is bounded away from zero. As in the no noise case, we choose the following score functions up to some proper normalizations.

$$\begin{aligned}\Psi_n &= (\Psi_n^1(\omega,\theta), \Psi_n^2(\omega,\theta))' = (-\frac{1}{\sqrt{n}}\frac{\partial l(a^2,\sigma^2)}{\partial \sigma^2}, -\frac{1}{n}\frac{\partial l(a^2,\sigma^2)}{\partial a^2})' \\ &= \Big(\frac{1}{2\sqrt{n}}\{\frac{\partial \log(\det\Omega)}{\partial\sigma^2} + Y'\frac{\partial\Omega^{-1}}{\partial\sigma^2}Y\}, \frac{1}{2n}\{\frac{\partial \log(\det\Omega)}{\partial a^2} + Y'\frac{\partial\Omega^{-1}}{\partial a^2}Y\}\Big)'\end{aligned}$$

and correspondingly,

$$\begin{aligned}\bar{\Psi}_n &= (\bar{\Psi}_n^1(\omega,\theta), \bar{\Psi}_n^2(\omega,\theta))' \\ &= \Big(\frac{1}{2\sqrt{n}}\{\frac{\partial \log(\det\Omega)}{\partial\sigma^2} + tr(\frac{\partial\Omega^{-1}}{\partial\sigma^2}\Sigma_0)\}, \frac{1}{2n}\{\frac{\partial \log(\det\Omega)}{\partial a^2} + tr(\frac{\partial\Omega^{-1}}{\partial a^2}\Sigma_0)\}\Big)'\end{aligned}$$

where Σ_0 is given by

$$\Sigma_0 = \begin{pmatrix} \int_0^{\tau_1}\sigma_t^2 dt + 2a_0^2 & -a_0^2 & 0 & \cdots & 0 \\ -a_0^2 & \int_{\tau_1}^{\tau_2}\sigma_t^2 dt + 2a_0^2 & -a_0^2 & \ddots & \vdots \\ 0 & -a_0^2 & \int_{\tau_2}^{\tau_3}\sigma_t^2 dt + 2a_0^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -a_0^2 \\ 0 & \cdots & 0 & -a_0^2 & \int_{\tau_{n-1}}^{T}\sigma_t^2 dt + 2a_0^2 \end{pmatrix}$$

Being aware that the convergence rates may be different for the two parameters, as shown in Section 2, we choose different normalizations accordingly. These normalizations are essential to ensure the identifiability condition.

Denote $\Omega^{-1} = (\omega^{ij})$, and $\epsilon_j = U_{\tau_j} - U_{\tau_{j-1}}$. The difference between Ψ_n^1 and $\bar{\Psi}_n^1$, for instance is

$$
\begin{aligned}
& 2\sqrt{n}(\Psi_n^1 - \bar{\Psi}_n^1) = Y' \frac{\partial \Omega^{-1}}{\partial \sigma^2} Y - tr(\frac{\partial \Omega^{-1}}{\partial \sigma^2} \Sigma_0) \qquad (12) \\
= \;& \sum_{i=1}^{n} \frac{\partial \omega^{ii}}{\partial \sigma^2} \underbrace{\{(\int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t)^2 - \int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt\}}_{\text{martingale difference}} + \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{\partial \omega^{ij}}{\partial \sigma^2} \int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t \int_{\tau_{j-1}}^{\tau_j} \sigma_t dW_t \\
+ \;& 2 \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial \omega^{ij}}{\partial \sigma^2} \epsilon_j \int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t + \underbrace{\sum_{i=1}^{n} \sum_{j \neq i, i-1, i+1}^{n} \frac{\partial \omega^{ij}}{\partial \sigma^2} \epsilon_i \epsilon_j}_{\text{due to the noise}} \\
+ \;& \underbrace{\sum_{i=1}^{n} \frac{\partial \omega^{ii}}{\partial \sigma^2} (\epsilon_i^2 - 2a_0^2) + \sum_{i=1}^{n} \frac{\partial \omega^{i,i-1}}{\partial \sigma^2} (\epsilon_i \epsilon_{i-1} + a_0^2) + \sum_{i=1}^{n} \frac{\partial \omega^{i,i+1}}{\partial \sigma^2} (\epsilon_i \epsilon_{i+1} + a_0^2)}_{\text{due to the noise}}
\end{aligned}
$$

The first term is a linear combination of martingale differences, while the second term is the sum of cross products over different pieces of integrals. The third one is a mixture of the martingale part and the noise part. The rest are purely related to the noise. It is clear that the first two terms as a whole, the third one and the rest are pairwise uncorrelated given the proposed assumptions.

Now we proceed with variance calculations. The lesson learned from the failure of RV indicates that the variance of the noise may dominate the others. So for the noise part, we accurately compute the variance of the derivatives with respect to different parameters.

Lemma 1. *Given Assumptions 1 and 2, we have*

$$\sum_{i=1}^{n} \omega^{ii} \{(\int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t)^2 - \int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt\} = O_p(1) \tag{13}$$

$$\sum_{i=1}^{n} \sum_{j \neq i}^{n} \omega^{ij} \int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t \int_{\tau_{j-1}}^{\tau_j} \sigma_t dW_t = O_p(n^{\frac{1}{4}}) \tag{14}$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \omega^{ij} \epsilon_j \int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t = O_p(n^{\frac{1}{4}}) \tag{15}$$

Lemma 2. *As to the noise part, we have*

$$var(\epsilon' \frac{\partial \Omega}{\partial \sigma^2} \epsilon) = \frac{\sqrt{T} a_0^4 n^{\frac{1}{2}}}{16 a^5 \sigma^3} + o(n^{\frac{1}{2}}) \tag{16}$$

$$var(\epsilon' \frac{\partial \Omega}{\partial a^2} \epsilon) = \frac{n(2a_0^4 + cum_4[U])}{a^8} + o(n) \tag{17}$$

By verifying the identifiability condition, solving the equations $\bar{\Psi}_n = 0$ and applying the above lemmas, we can prove

Theorem 4. Ψ_n *and* $\bar{\Psi}_n$ *are as given,* $\hat{\theta}_n = (\hat{\sigma}_n^2, \hat{a}_n^2)$ *is as defined in Theorem 2. Given Assumptions 1 and 2, the QMLE*$(\hat{\sigma}^2, \hat{a}^2)$ *satisfies:* $\hat{a}_n^2 - a_0^2 \xrightarrow{\mathrm{P}} 0$ *and* $\hat{\sigma}_n^2 - \frac{1}{T}\int_0^T \sigma_t^2 dt \xrightarrow{\mathrm{P}} 0.$

As expected, the consistency of the volatility estimator still holds, even though the volatility process becomes stochastic.

4.2 Central Limit Theorem of the QMLE

Although we have shown the consistency of the QMLE, we have not explored whether the QMLE has the optimal convergence rates, neither have we shown anything about the magnitude of the asymptotic variance or possible asymptotic bias. To answer these questions, the following lemma and theorem provide the central limit theory.

Lemma 3. *Given Assumptions 1 and 2, we have*[4]

$$\begin{pmatrix} n^{\frac{1}{4}}(\Psi_n^1 - \bar{\Psi}_n^1) \\ n^{\frac{1}{2}}(\Psi_n^2 - \bar{\Psi}_n^2) \end{pmatrix} \xrightarrow{\mathcal{L}_{\mathrm{X}}} MN\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{4}(\frac{5\int_0^T \sigma_t^4 dt}{16 a \sigma^7 \sqrt{T}} + \frac{a_0^2 \int_0^T \sigma_t^2 dt}{8\sigma^5 a^3 \sqrt{T}} + \frac{a_0^4 \sqrt{T}}{16 a^5 \sigma^3}) & 0 \\ 0 & \frac{2a_0^4 + cum_4[U]}{4a^8} \end{pmatrix} \Big)$$

Theorem 5. Ψ_n *and* $\bar{\Psi}_n$ *are as given,* $\hat{\theta}_n = (\hat{\sigma}_n^2, \hat{a}_n^2)$ *is as defined in Theorem 2. Given Assumptions 1 and 2, the QMLE* $(\hat{\sigma}^2, \hat{a}^2)$ *satisfies:*

$$\begin{pmatrix} n^{\frac{1}{4}}(\hat{\sigma}^2 - \frac{1}{T}\int_0^T \sigma_t^2 dt) \\ n^{\frac{1}{2}}(\hat{a}^2 - a_0^2) \end{pmatrix} \xrightarrow{\mathcal{L}_{\mathrm{X}}} MN\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{5a_0 \int_0^T \sigma_t^4 dt}{T(\int_0^T \sigma_t^2 dt)^{\frac{1}{2}}} + \frac{3(\int_0^T \sigma_t^2 dt)^{\frac{3}{2}} a_0}{T^2} & 0 \\ 0 & 2a_0^4 + cum_4[U] \end{pmatrix} \Big)$$

In the constant volatility case, QMLE attains the optimal efficiency, as can be expected since the QMLE is constructed in the same way as the MLE. In general, QMLE achieves the optimal convergence rates. However, this QMLE method might not provide a straightforward estimator for the integrated quarticity, that is, $\int_0^T \sigma_t^4 dt$. Thus, as to the construction of confidence intervals, one can instead use the method given by Jacod et al. (2007).

[4] MN is a notation of mixed normal used in Barndorff-Nielsen et al. (2008), and $\mathcal{L}_{\mathrm{X}}$ means $\sigma(X)$-stable convergence in law.

4.3 Robustness of the QMLE

4.3.1 Drift

What happens to our conclusions if the underlying X process has a drift? More precisely, suppose X_t to be of the Itô type,

$$dX_t = \mu_t dt + \sigma_t dW_t$$

where μ_t is a locally bounded and progressively measurable process. Instead of parameterizing the drift term and modifying our likelihood estimator accordingly, we completely ignore the presence of drift, or in other words, misspecify the model with drift 0. In this case, the estimator is unchanged, and we only need to replace $\int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t$ with $\int_{\tau_{i-1}}^{\tau_i} \mu_t dt + \int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t$ in the proof. Loosely speaking, this will not alter our conclusions since in the high frequency setting, the drift is of order dt, which is mathematically negligible with respect to the diffusive component of order $(dt)^{\frac{1}{2}}$. An alternative argument given by Mykland and Zhang (2009) is to zero out the drift by changing probability measures.

4.3.2 Random Sampling Intervals

What if the observations are made randomly? If the sampling intervals between two consecutive observations are i.i.d., and independent of the price process, we may add one more misspecification that the data are regularly spaced; hence we obtain the same estimator as before. Consistency can be established directly by conditioning on the observation time. In fact, this estimator can be viewed as the Pretend Fixed Maximum Likelihood (PFML) estimator discussed in Aït-Sahalia and Mykland (2003). In light of this, we may perform the Full Information Maximum Likelihood (FIML) or Integrated Out Maximum Likelihood (IOML) estimation, in order to utilize the information regarding the distribution of the random intervals, if available.

Also, it is possible to extend the i.i.d. sampling scheme to endogenous and stochastically spaced observations, by considering the time-changed processes as in Barndorff-Nielsen et al. (2008). The extension is effortless in light of the connection between QMLE and RKs spelled out in Section 5.

4.3.3 Non-Gaussian and Serial Dependent Microstructure Noise

What about the robustness with respect to the noise? From Lemma 2 and Theorem 4, we find that the distribution of the i.i.d. noise does not affect the consistency. In other words, whatever the distribution of the noise is, it is misspecified to be Gaussian, and the QMLE obtained by maximizing (2) gives the same estimator of a_0^2 and the same order of convergence rate, though the asymptotic variance may be different.

On the other hand, if the noises are in fact time-dependent, we can combine the QMLE with the subsampling method. For instance, if the noise itself features an MA(k) structure, we can divide the whole sample into $k+1$ disjoint parts such that the noises associated with the adjacent points in each subsample are uncorrelated. Then, we can simply apply the QMLE to each subsample and aggregate the estimates by taking the average.

4.3.4 Jumps

If the price process has jumps, then $\hat{\sigma}^2$ will converge to $\frac{1}{T}(\int_0^T \sigma_t^2 dt + \sum_t (\Delta X_t)^2)$ instead, which coincides with the TSRV estimator. The problem of separating jumps from volatility in this setting is more tedious, and may contribute little, since in most cases large jumps do not happen very often within a day. If they do occur, we may use the wavelet method in Fan and Wang (2007) to remove jumps before estimation, or separate the estimation periods by jumps and add up every piece of integrated volatility together, if the positions of jumps can be located accurately.

5 Comparisons with Realized Kernels

5.1 Estimation Methods

Realized Kernels (RKs) include a series of nonparametric estimators designed for volatility estimation in the presence of noise. Flat-top RKs take on the following form

$$K(\tilde{X}_\tau) = \gamma_0(\tilde{X}_\tau) + \sum_{h=1}^{n-1} k(\frac{h-1}{H})(\gamma_h(\tilde{X}_\tau) + \gamma_{-h}(\tilde{X}_\tau)) \tag{18}$$

where

$$\gamma_h(\tilde{X}_\tau) = \sum_{j=1}^{n} (\tilde{X}_{\tau_j} - \tilde{X}_{\tau_{j-1}})(\tilde{X}_{\tau_{j-h}} - \tilde{X}_{\tau_{j-h-1}}) \tag{19}$$

is the h^{th} sample autocovariance function and the kernel $k(\cdot)$ is a weight function. The most common finite-lag flap-top kernels are of the modified Tukey-Hanning type, defined by:

$$k_{TH_p}(x) = \sin^2(\frac{\pi}{2}(1-x)^p) \cdot 1_{\{0 \le x \le 1\}}$$

In addition, there are infinite-lag realized kernels which may assign nonzero weight to all autocovariances, such as the optimal kernel:

$$k_{opt}(x) = (1+x)e^{-x}$$

As expected, the statistical properties of RKs vary as different kernels and bandwidths are selected. Therefore, it enables us to make different choices for specific purposes. However, the choice of the bandwidth may also become a burden in practice, in that the optimal bandwidth given by the theory cannot be estimated very accurately, and that the rule-of-thumb approximation of the bandwidth may not perform as well as the one selected in ad hoc ways. This problem may become even worse if the estimates were sensitive to the choice of bandwidth. By contrast, the QMLE is designed in a parametric way, which is free of bandwidth and kernel selection, while sharing the desired model-free feature with nonparametric estimators. It is by nature a different estimator and may not be regarded as one of the RKs directly.

5.2 Asymptotic Behavior and Finite Sample Performance

In regard to asymptotic efficiency, the kernel and bandwidth can be chosen for RKs to achieve the same optimal convergence rate as the QMLE. Nevertheless, when volatility is constant, the asymptotic variance of finite-lag kernels can only approximate the parametric variance bound, which, by contrast, can be obtained by the QMLE and the optimal kernel with infinite lags. When volatility is stochastic, the relative efficiency of the QMLE and RKs depends on the extent of heteroskedasticity. More precisely, we have[5]

$$\frac{Avar(\text{RK})}{Avar(\text{QMLE})} = \frac{16\sqrt{\rho k_0 k_1}}{3} \frac{\left(1+\sqrt{1+3k_0k_2/\rho k_1^2}\right)^{-\frac{1}{2}} + \left(1+\sqrt{1+3k_0k_2/\rho k_1^2}\right)^{\frac{1}{2}}}{5\rho^{-\frac{1}{2}} + 3\rho^{\frac{3}{2}}}$$

where $\rho = \int_0^T \sigma_u^2 du / \sqrt{T \int_0^T \sigma_u^4 du}$ measures the variability of the volatility process, and k_0, k_1 and k_2 are constants determined by the selected kernel. Figure 1 plots the relative efficiency of

[5] The asymptotic variance of the RK obtained here requires a refinement of the endpoints as well as the optimal bandwidth.

four typical RKs against the QMLE. Apparently, the QMLE tends to be more favorable than finite-lag kernels as ρ becomes larger, while RKs are better when ρ is small. The optimal kernel with theoretically optimal bandwidth, however, fully dominates the QMLE asymptotically except for the case $\rho = 1$, when volatility is constant. Intuitively, the smaller ρ is, the further the misspecified model deviates from the truth.

A major drawback of RKs is that they require a number of out-of-period intraday returns because of the construction of the autocovariance estimator $\gamma_h(\tilde{X}_\tau)$. For this reason, infinite-lag kernels, in particular, are not implementable empirically. In practice, the feasible autocovariance estimator is implemented by

$$\tilde{\gamma}_{\pm h}(\tilde{X}_\tau) = \sum_{j=H+1}^{n-H} (\tilde{X}_{\tau_j} - \tilde{X}_{\tau_{j-1}})(\tilde{X}_{\tau_{j-h}} - \tilde{X}_{\tau_{j-h-1}}) \tag{20}$$

With finite-lag flap-top kernels, the cut-off near the boundary is not an issue asymptotically, however it may yield a large bias in finite sample. As the sample size n decreases, namely, H/n increases, this bias becomes more evident. The next section applies Monte Carlo simulations to demonstrate the edge effect.

5.3 Quadratic Representation and Weighting Matrices

An intuitive way to understand the similarities and differences between the QMLE and RKs is to regard them as quadratic estimators. More specifically, we have the following quadratic iterative representation of the QMLE.

Theorem 6. *The QMLE $(\hat{\sigma}^2, \hat{a}^2)$ satisfies the following equations:*

$$\hat{\sigma}^2 T = Y' W_1(\hat{\sigma}^2, \hat{a}^2) Y \tag{21}$$

$$\hat{a}^2 = Y' W_2(\hat{\sigma}^2, \hat{a}^2) Y \tag{22}$$

The weighting matrices satisfy:

$$W_1(\sigma^2, a^2) = \frac{n \cdot tr(\Omega^{-2}\Lambda) \cdot \Omega^{-1}\Lambda\Omega^{-1} - n \cdot tr(\Omega^{-2}\Lambda^2) \cdot \Omega^{-2}}{(tr(\Omega^{-2}\Lambda))^2 - tr(\Omega^{-2}) \cdot tr(\Omega^{-2}\Lambda^2)} \tag{23}$$

$$W_2(\sigma^2, a^2) = \frac{tr(\Omega^{-2}\Lambda) \cdot \Omega^{-2} - tr(\Omega^{-2}) \cdot \Omega^{-1}\Lambda\Omega^{-1}}{(tr(\Omega^{-2}\Lambda))^2 - tr(\Omega^{-2}) \cdot tr(\Omega^{-2}\Lambda^2)} \tag{24}$$

where Ω is given by (3), and

$$\Lambda = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & -1 & 2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix}$$

Also, $W_1(\sigma^2, a^2)$ and $W_2(\sigma^2, a^2)$ depend on σ^2 and a^2 only through $\lambda^2 = a^2/\sigma^2 T$.

Similarly, a feasible realized kernel estimator can be expressed as

$$K(\tilde{X}_\tau) = Y'WY \tag{25}$$

where W is determined by the kernel $k(\cdot)$ and bandwidth H.

$$\begin{aligned} W_{i,i} &= 1_{\{1+H \leq i \leq n-H\}} \\ W_{i,j} &= k(\frac{|i-j|-1}{H}) \cdot 1_{\{1 \leq |i-j| \leq H\}} \cdot 1_{\{1+H \leq j \leq n-H\}} \end{aligned}$$

Figure 2 plots the weighting matrices against the row and column indices for the QMLE and Turkey-Hanning$_2$ kernel. The plot displays similar features such as the weights on the diagonals are either 1 or very close to 1 within several lags, and decay to zero beyond that. The following theorem illustrates the implicit connection between the two estimators in terms of the asymptotic behavior.

Theorem 7. *The QMLE is asymptotically equivalent to the optimal kernel with implicit bandwidth $H = \hat{\lambda} \cdot n^{\frac{1}{2}} = a_0/(\int_0^T \sigma_t^2 dt)^{\frac{1}{2}} n^{\frac{1}{2}}$. In other words, for any $K = n^{\frac{1}{2}+\delta}$, $0 < \delta < \frac{1}{2}$, and any $K \leq i, j \leq n - K$, we have*

$$W_{1,i,j}(\sigma^2, a^2) \approx k_{opt}(\frac{|i-j|}{\lambda \cdot n^{\frac{1}{2}}})$$

Theorem 7 also points out that the weighting matrix $W_1(\sigma^2, a^2)$ is approximately a symmetric Toeplitz matrix with equal weight along the diagonal, barring the boundary effect. The implicit bandwidth depends on the parameters of interest and is suboptimal. It therefore explains why the optimal kernel with optimal bandwidth asymptotically dominates the QMLE in Figure 1. Nevertheless, the ending points of the diagonals have different patterns, indicating different treatments of the edge effect. Apparently, the QMLE controls the weights on the boundary in a more natural way, and hence has better finite sample performance.

The quadratic representation also sheds light on the differences in the procedures of estimation. The bandwidth H of the RK is either estimated as first step or selected in an ad hoc way, while the bandwidth $\hat{\lambda} \cdot n^{\frac{1}{2}}$ of the QMLE is automatically updated by the optimization algorithm, or more intuitively, by iterating (21) and (22). The weighting matrix of the QMLE is therefore more adaptive. In view of this, it is natural to construct a one-step alternative for the QMLE, which, instead of running nonlinear optimization, employs a consistent plug-in of $\hat{\lambda}$ for (21) and (22). This one-step volatility estimator coincides with one of the unbiased quadratic estimators given by Sun (2006), where its asymptotic properties are discussed under the constant volatility assumption. The quadratic forms of other nonparametric estimators are also included in Andersen et al. (2009).

6 Simulation Studies with High Frequency Data

6.1 Asymptotic Behavior of the QMLE

We at first conduct Monte Carlo simulations to verify the asymptotic results given in Theorem 5. We fix T as 1 day. Within $[0, T]$, the data are simulated using Euler scheme based on stochastic volatility models, for instance the Heston Model with jumps in volatility process.

$$
\begin{aligned}
dX_t &= \mu dt + \sigma_{t-} dW_t \\
d\sigma_t^2 &= \kappa(\bar{\sigma}^2 - \sigma_t^2)dt + \delta\sigma_{t-} dB_t + \sigma_{t-} J_t^V dN_{2t}
\end{aligned}
$$

where $E(dW_t \cdot dB_t) = \rho dt$. The arrival of transactions follows a Poisson process and the mean interarrival time is 1 second. The true values of parameters are consistent with those chosen by Aït-Sahalia and Yu (2009). Specifically, the drift μ is 0.03, and to add the leverage effect, ρ is selected to be -0.75. $\kappa = 5$ and the volatility of volatility is $\delta = 0.4$. σ_0^2 is sampled from the stationary distribution of the CIR process, that is Gamma$(2\kappa\bar{\sigma}^2/\delta^2, \delta^2/2\kappa)$, so that the unconditional mean of volatility process is exactly $\bar{\sigma}^2$, chosen as 0.1. The noise satisfies normal distribution with standard deviation $a_0 = 0.5\%$. The jumps follow a Poisson process N_{2t} independent of the price and volatility processes with intensity $\lambda = 12$. The jump size in volatility is $J_t^V = \exp(z)$, where $z \sim N(-5, 1)$. The number of Monte Carlo sample paths is

10000.

Figure 3 shows the histograms of the standardized estimates compared with their asymptotic distributions. Moreover, Table 1 compares the sample quartile statistics with their theoretic benchmarks given by the standard Gaussian distribution. All these simulation results reconfirm our asymptotic theory.

6.2 Comparisons with RKs

For comparison purpose, we implement the Tukey-Hanning$_2$ kernel which is one of the flap-top kernels that converge at the best rate. It is considerably efficient compared with other kernels in Barndorff-Nielsen et al. (2008), and it does not require too many out-of-period data.

As a benchmark, we first implement the Tukey-Hanning$_2$ kernel with the infeasible bandwidth, $H = c^* \xi n^{\frac{1}{2}}$ with c^* given by

$$c^* = \sqrt{\rho \frac{k_1}{k_0}(1 + \sqrt{1 + \frac{3k_0 k_2}{\rho k_1^2}})}$$

where $\xi^2 = a_0^2 / \sqrt{T \int_0^T \sigma_u^4 du}$, $\rho = \int_0^T \sigma_u^2 du / \sqrt{T \int_0^T \sigma_u^4 du}$, $k_0 = 0.219$, $k_1 = 1.71$ and $k_2 = 41.7$. Also, the edge effect is eliminated in that the calculation, using formula (19), includes the out-of-period data. Hence, this estimator would provide us with the best Tukey-Hanning$_2$ kernel estimate (RK$_1$) in theory.

Next, we make the edge effect stand out by approximating the autocovariances using (20), while keeping the theoretical optimal bandwidth as before. We denote the second estimator as RK$_2$.

Last, to mimic the empirical applications of these estimators, we redo the experiments with the simple rule-of-thumb choice of the bandwidth:

$$\hat{H} = 5.74 \hat{a} \sqrt{n / RV_{10}(\tilde{X})}$$

where $\hat{a} = \sqrt{RV(\tilde{X})/2n}$ and $RV_{10}(\tilde{X})$ is the RV estimator based on 10-min returns. So we obtain the corresponding RK$_3$ and RK$_4$.

The simulations are based on the same stochastic volatility model and parameters as before with a longer time window $[0, 3T]$. We regard $[T, 2T]$ as in-sample period, and the rest time interval is only used for infeasible kernels. The results of estimation are reported in Table 2. It is evident that the QMLE dominates the four implementations of the Tukey-Hanning$_2$ kernel in terms of the RMSE. As expected, RK_1 is very close to the QMLE, and better than the other three kernels, in that it employs more data and is equipped with theoretically optimal bandwidth. Comparing the first two kernels, we can also see that the edge effect results in a large bias when the sample size is relatively small, and become dampened as the sample size goes up. In addition, the differences between the first two kernels and the last two are the losses due to the suboptimal choice of the bandwidth. In practice, only RK_4 can be applied, which may suffer from larger bias and variance because of the two sources of losses.

7 Empirical Work with the Euro/US Dollar Future Prices

In this section, we are interested in estimating the realized variance of the Euro/US dollar futures carried out on the Chicago Mercantile Exchange (CME) in the year 2008. The contracts are actively traded on the 24-hour clock, and quoted in terms of the unit value of the Euro as measured in US dollars. The high frequency data are available from Tick Data Inc.

The foreign exchange markets are less active during the weekends and holidays; therefore, we eliminate the transactions on Saturdays and Sundays as well as US federal holidays. In addition, we also exclude January 2, the day after Thanksgiving, December 24 to 26, and December 31. Last, we make everyday begin at 5pm Chicago Time when the electronic trading starts so as to eliminate the potential price jumps between the one hour trading gap from 4pm to 5pm. The summary statistics of the data are provided in Table 3. Evidently, the microstructure noise displays an MA(1) structure.

We apply both the QMLE and the Tukey-Hanning$_2$ kernel with the rule-of-thumb bandwidth and plot the annualized daily realized volatility in Figure 4. It is apparent from the plot that the two methods give almost identical estimates and the differences are statistically insignificant. This is in agreement with our Monte Carlo simulation result, where the two estimates are indistinguishable when the trading frequency is as high as several seconds.

Moreover, Figure 4 provides ample evidence of structure breaks or jumps in the volatility process starting from September 2008, which echoes the fact that, since then, the global financial crisis has entered its most critical stage. Further, these jumps or large movements are usually associated with financial news: for instance, the government seizure of Fannie Mae and Freddie Mac, Lehman Brothers' bankruptcy in early September, and the collapse of three major banks of Iceland in early October, etc. These findings are harder to obtain by modeling lower frequency data.

8 Conclusions

This article contributes to the estimation of integrated volatility by showing that the popular MLE, as a new quasi-estimator, is consistent, efficient and robust with respect to stochastic volatility. Moreover, this parametric estimator only involves an optimization procedure, which is free from bandwidth selection, and hence very convenient in practice. More interestingly, this seemingly inappropriate estimator turns out to be asymptotically equivalent to the optimal realized kernel with an implicitly specified bandwidth, and dominant over alternative realized kernels in terms of the finite sample accuracy.

This article makes another contribution to the analysis of model specification by extending the classic asymptotic theory of the QMLE to a stochastic parameter setting, and by showing an example of misspecifying models on purpose, which gives rise to facility and feasibility in estimation. This study may be applicable to more cases where stochastic volatility plagues the estimation, such as covariance estimation, or more generally, where the object of interest is stochastic.

References

Aït-Sahalia, Y., Mykland, P. and Zhang, L., 2005. How often to sample a continuous-time process in the presence of market microstructure noise. Review of Financial Studies 18, 351-416.

Aït-Sahalia, Y., Mykland, P. and Zhang, L., 2006. Ultra high frequency volatility estimation with dependent microstructure noise. Forthcoming in the Journal of Econometrics.

Aït-Sahalia, Y. and Mykland, P., 2003. The effect of random and discrete sampling when estimating continuous time diffusions. Econometrica 71, 483-549.

Aït-Sahalia, Y. and Yu, J., 2009. High frequency market microstructure noise estimates and liquidity measures. Annals of Applied Statistics 3, 422-457.

Amemiya, T., 1973. Regression analysis when the dependent variable is truncated normal. Econometrica 41, 997-1016.

Andersen, T.G., Bollerslev, T. and Meddahi, N., 2009. Realized volatility forecasting and market microstructure noise. Forthcoming in the Journal of Econometrics.

Andersen, P.K. and Gill, R.D., 1982. Cox's regression model for counting processes: a large sample study. Annals of Statistics 19, 1100-1120.

Bandi, M. and Russell, J., 2008. Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations. Discussion paper, University of Chicago, Graduate School of Business.

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

Barndorff-Nielsen, O.E., and Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of Royal statistical Society B 64, 253-280.

Bates, C. and White, H., 1985. A unified theory of consistent estimation for parametric models. Econometric Theory 1, 151-178.

Delattre, S. and Jacod, J., 1997. A central limit theorem for normalized functions of the increments of a diffusion process in the presence of round-off errors. Bernoulli 3, 1-28.

Domowitz, I. and White, H., 1982. Misspecified models with dependent observations. Journal of Econonmetrics 20, 35-58.

Fan, J. and Wang, Y., 2007. Multi-Scale jump and volatility analysis for high-frequency financial data. Journal of the American Statistical Association 102, 1349-1362.

Gatheral, J. and Oomen, R., 2007. Zero-Intelligence Realized Variance Estimation. Working Paper.

Gloter, A. and Jacod, J., 2001. Diffusions with measurement errors: I. Local asymptotic normality. ESAIM 5, 225-242.

Hamilton, J.D., 1994. Time Series Analysis, Princeton University Press.

Hansen, P. R., and Lunde, A., 2006. Realized variance and market microstructure noise. Journal of Business and Economic Statistics 24, 127-218.

Hansen, P. R., Large, J., and Lunde, A., 2008. Moving average-based estimators of integrated variance. Econometrics Review 27, 79-111.

Jacod, J., 1994. Limit of random measures associated with the increments of a Brownian semmimartingale. Discussing paper.

Jacod, J., 2007. Statistics and high frequency data. Unpublished lecture notes, Séminaire Européen de Statistique 2007: Statistics for stochastic differential equations models.

Jacod, J., 2008. Asymptotic properties of realized power variations and related functionals of semimartingales. Stochastic Processes and their Applications 118, 517C559.

Jacod, J., Li, Y., Mykland, P., Podolskij, M., and Vetter, M., 2007. Microstructure noise in the continuous case: the pre-averaging approach. Working paper.

Jacod, J. and Shiryaev, A.N., 2003. Limit theorems for stochastic processes, 2nd Edition. New York: Springer-Verlag.

Kalnina, I. and Linton, O., 2008. Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error. Journal of Econometrics, forthcoming.

Karatzas, I. and Shreve, S., 1991. Brownian motion and stochastic calculus, 2nd Edition, GTM 113, Springer-Verlag Press.

Kullbacks, S. and Leibler, R., 1951. On information and sufficiency. Annals of Mathematical Statistics 22, 79-86.

Li, Y and Mykland, P., 2007. Are Volatility Estimators Robust with Respect to Modeling Assumptions? Bernoulli 13, 601-622.

McCullagh, P., 1987. Tensor Methods in Statistics. Chapman and Hall, London, U.K.

Mykland, P. and Zhang, L., 2006. ANOVA for diffusions and Itô processes. Annals of Statistics 34, 1931-1963.

Mykland, P.A., and Zhang, L., 2009. Inference for continuous semimartingales observed at high frequency. Econometrica 77, 1403 - 1445.

Newey, W., 1991. Uniform convergence in probability and stochastic equicontinuity, Econometrica 59, 1161-1167.

Newey, W. and McFadden, D., 1994. Large sample estimation and hypothesis testing, Handbook of Econometrics, Vol IV, Chapter 36.

Rockafellar, T., 1970. Convex Analysis, Princeton University Press.

Sun, Y., 2006. Best quadratic unbiased estimators of integrated variance in the presence of market microstructure noise. Working paper.

Van Der Vaart, A.W. 2000. Asymptotic Statistics, Cambridge University Press.

White, H., 1980. Nonlinear regression on cross-section data. Econometrica 48, 721-746.

White, H., 1982. Maximum likelihood estimation of misspecified models. Econometrica 50, 1-26.

Zhang, L., Mykland, P.A. and Aït-Sahalia, Y., 2005. A tale of two time scales: Determining integrated volatility with noisy high frequency data. Journal of the American Statistical Association 100, 1394-1411.

Zhang, L., 2006. Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. Bernoulli 12, 1019-1043.

Zhou, B., 1996. High-frequency data and volatility in foreign-exchange rates. Journal of Business & Economic Statistics 14, 45-52.

Appendix

1 Proof of Theorem 1

For each $\delta > 0$, with probability approaching 1 (w.p.a.1), we have[6]

$$\begin{aligned} &Q_n(\omega, \hat{\theta}_n) > Q_n(\omega, \theta_n^*) - \tfrac{\delta}{3} \\ &\left.\begin{aligned} &\bar{Q}_n(\omega, \hat{\theta}_n) > Q_n(\omega, \hat{\theta}_n) - \tfrac{\delta}{3} \\ &Q_n(\omega, \theta_n^*) > \bar{Q}_n(\omega, \theta_n^*) - \tfrac{\delta}{3} \end{aligned}\right\} \text{Due to Uniform Convergence} \end{aligned}$$

Thus, $\forall\ \delta > 0$, w.p.a.1, $\bar{Q}_n(\omega, \hat{\theta}_n) > \bar{Q}_n(\omega, \theta_n^*) - \delta$.

Further, for any $\epsilon > 0$, let $N(\omega) := \{\theta : \|\theta - \theta_n^*\| < \epsilon\}$. Then for each ω, $N^c \cap \Theta$ is compact, and $\max\limits_{\theta \in N^c \cap \Theta} \bar{Q}_n(\omega, \theta) = \bar{Q}_n(\omega, \tilde{\theta}) \leq \bar{Q}_n(\omega, \theta_n^*)$. Let

$$\delta_n(\omega) := \bar{Q}_n(\omega, \theta_n^*) - \max_{\theta \in N^c \cap \Theta} \bar{Q}_n(\omega, \theta)$$

We have, for any $\delta < \delta_0$,

$$\begin{aligned} & P(\bar{Q}_n(\omega, \hat{\theta}_n) > \max_{\theta \in N^c \cap \Theta} \bar{Q}_n(\omega, \theta)) \\ = \ & P(\bar{Q}_n(\omega, \hat{\theta}_n) > \bar{Q}_n(\omega, \theta_n^*) - \delta_n) \\ \geq \ & P(\bar{Q}_n(\omega, \hat{\theta}_n) > \bar{Q}_n(\omega, \theta_n^*) - \delta_n, \delta_n > \delta) \\ \geq \ & P(\bar{Q}_n(\omega, \hat{\theta}_n) > \bar{Q}_n(\omega, \theta_n^*) - \delta, \delta_n > \delta) \to 1 \end{aligned}$$

Therefore, $P(\|\hat{\theta}_n - \theta_n^*\| < \epsilon) \to 1$.

2 Proof of Theorem 2

The proof is similar to Van Der Vaart (2000)'s proof of consistency of the M-estimators. It follows from Theorem 1, on applying it to $Q_n(\omega, \theta) = -\|\Psi_n(\omega, \theta)\|$ and $\bar{Q}_n(\omega, \theta) = -\|\bar{\Psi}_n(\omega, \theta)\|$.

[6] The statements in this proof are well posed if any desired measurability is guaranteed. However, this measurability issue can be ignored by redefining all concepts in terms of outer measure, see Newey and McFadden (1994, pp 2121).

3 Proof of Theorem 3

For simplicity, it is sufficient to prove this result for $k = 1$. Because

$$\Psi_n(\omega, \theta) - \Psi_n(\omega, \theta_n^*) = \nabla \Psi_n(\omega, \tilde{\theta}_n)(\theta - \theta_n^*)$$

Plug in $\hat{\theta}_n$ and multiply both sides by V_n, then we have

$$V_n(\omega) \nabla \Psi_n(\omega, \tilde{\theta}_n)(\hat{\theta}_n - \theta_n^*) = -V_n(\omega) \Psi_n(\omega, \theta_n^*)$$

Since $\hat{\theta}_n - \theta_n^* \xrightarrow{\mathrm{P}} 0$, $\nabla \bar{\Psi}_n$ is stochastic equicontinuous, and $|\nabla \Psi_n(\omega, \theta) - \nabla \bar{\Psi}_n(\omega, \theta)| \xrightarrow{\mathrm{P}} 0$, uniformly for all $\theta \in \Theta$, it follows from an analogous reasoning as in Theorem 2.3 in Domowitz and White (1982) that $\nabla \Psi_n(\omega, \tilde{\theta}_n) - \nabla \bar{\Psi}_n(\omega, \theta_n^*) \xrightarrow{\mathrm{P}} 0$, which concludes the proof.

4 Proof of Lemma 1

Let $\Omega^{-1} = (\omega^{ij})$ and $X_{\tau_i} = \int_0^{\tau_i} \sigma_t dW_t$. We also define $M_{\tau_i} = X_{\tau_i}^2 - \langle X_{\tau_i}, X_{\tau_i} \rangle = (\int_0^{\tau_i} \sigma_t dW_t)^2 - \int_0^{\tau_i} \sigma_t^2 dt$. Then $\{X_{\tau_i}\}_{1 \le i \le n}$ and $\{M_{\tau_i}\}_{1 \le i \le n}$ are martingales. Also, we can add one more condition that $B^{-1} \le \sigma_t^2 \le B$, $\forall\, t \in [0, T]$, in that a regular localization procedure always applies. For convenience, we change the variables:

$$\gamma^2(1 + \eta^2) = \sigma^2 \Delta + 2a^2$$
$$\gamma^2 \eta = -a^2$$

Then the inverse change of variables is given by,

$$\eta = \frac{1}{2a^2}\{-2a^2 - \sigma^2\Delta + \sqrt{\sigma^2\Delta(4a^2 + \sigma^2\Delta)}\} \tag{26}$$
$$\gamma^2 = \frac{1}{2}\{2a^2 + \sigma^2\Delta + \sqrt{\sigma^2\Delta(4a^2 + \sigma^2\Delta)}\} \tag{27}$$

and we have

$$\omega^{ij} = \frac{(-\eta)^{|i-j|} - (-\eta)^{i+j} - (-\eta)^{2n-i-j+2} + (-\eta)^{2n-|i-j|+2}}{\gamma^2(1-\eta^2)(1-\eta^{2n+2})} \tag{28}$$

One can check $\omega^{ij} \ge 0$. In addition, ω^{ii} is a concave function of i, and clearly attains its maximum at $i = \frac{n+1}{2}$ and minimum at the boundary. In addition, by (26) and (27)

$$\eta \;\; = \;\; -1 + \sqrt{\frac{\sigma^2 T}{a^2}} n^{-\frac{1}{2}} - \frac{\sigma^2 T}{2a^2} n^{-1} + O(n^{-\frac{3}{2}})$$

$$\begin{aligned}
\eta^2 &= 1-2\sqrt{\frac{\sigma^2T}{a^2}}n^{-\frac{1}{2}}+\frac{2\sigma^2T}{a^2}n^{-1}+O(n^{-\frac{3}{2}})\\
\log\eta^2 &= -2\sqrt{\frac{\sigma^2T}{a^2}}n^{-\frac{1}{2}}+O(n^{-\frac{3}{2}})\\
\gamma^2 &= a^2+\sqrt{a^2\sigma^2T}n^{-\frac{1}{2}}+\frac{\sigma^2T}{2}n^{-1}+O(n^{-\frac{3}{2}})
\end{aligned}$$

Then, it can be easily shown that $\omega^M := \omega^{\frac{n+1}{2},\frac{n+1}{2}} = O(n^{\frac{1}{2}})$. By the martingale property, we have

$$\begin{aligned}
&E(\sum_{i=1}^{n}\omega^{ii}\{(\int_{\tau_{i-1}}^{\tau_i}\sigma_t dW_t)^2-\int_{\tau_{i-1}}^{\tau_i}\sigma_t^2dt\})^2\\
=\ &\sum_{i=1}^{n}(\omega^{ii})^2E\{(\int_{\tau_{i-1}}^{\tau_i}\sigma_t dW_t)^2-\int_{\tau_{i-1}}^{\tau_i}\sigma_t^2dt\}^2\\
\leq\ &(\omega^M)^2\sum_{i=1}^{n}E\{(\int_{\tau_{i-1}}^{\tau_i}\sigma_t dW_t)^2-\int_{\tau_{i-1}}^{\tau_i}\sigma_t^2dt\}^2
\end{aligned}$$

As in the no noise case, we have

$$\sum_{i=1}^{n}\{(\int_{\tau_{i-1}}^{\tau_i}\sigma_t dW_t)^2-\int_{\tau_{i-1}}^{\tau_i}\sigma_t^2dt\} = O_p(n^{-\frac{1}{2}}) \tag{29}$$

Combining the two, we obtain

$$E(\sum_{i=1}^{n}\omega^{ii}\{(\int_{\tau_{i-1}}^{\tau_i}\sigma_t dW_t)^2-\int_{\tau_{i-1}}^{\tau_i}\sigma_t^2dt\})^2 = (O(n^{\frac{1}{2}}))^2O(n^{-1}) = O(1)$$

Thus, we prove (13) by Chebyshev's inequality. To prove (14), note that $\omega^{ij}=\omega^{ji}$, so it follows from the martingale property that

$$\begin{aligned}
&E(\sum_{i=1}^{n}\sum_{j\neq i}^{n}\omega^{ij}\int_{\tau_{i-1}}^{\tau_i}\sigma_t dW_t\int_{\tau_{j-1}}^{\tau_j}\sigma_t dW_t)^2\\
=\ &4E(\sum_{i}\sum_{j>i}\sum_{k}\sum_{l>k}\omega^{ij}\omega^{kl}\int_{\tau_{i-1}}^{\tau_i}\sigma_t dW_t\int_{\tau_{j-1}}^{\tau_j}\sigma_t dW_t\int_{\tau_{k-1}}^{\tau_k}\sigma_t dW_t\int_{\tau_{l-1}}^{\tau_l}\sigma_t dW_t)\\
=\ &4\sum_{j=1}^{n}E\{(\int_{\tau_{j-1}}^{\tau_j}\sigma_t dW_t)^2(\sum_{k=1}^{j-1}\omega^{kj}\int_{\tau_{k-1}}^{\tau_k}\sigma_t dW_t)^2\}\\
\leq\ &4B\Delta\sum_{j=1}^{n}E(\sum_{k=1}^{j-1}\omega^{kj}\int_{\tau_{k-1}}^{\tau_k}\sigma_t dW_t)^2
\end{aligned}$$

$$
\begin{aligned}
&= 4B\Delta \sum_{j=1}^{n} \sum_{k=1}^{j-1} (\omega^{kj})^2 E(\int_{\tau_{k-1}}^{\tau_k} \sigma_t dW_t)^2 \\
&\leq 4B^2\Delta^2 \sum_{j=1}^{n} \sum_{k<j} (\omega^{kj})^2 \\
&= O(n^{\frac{1}{2}})
\end{aligned}
$$

Last, we prove (15). Because for any $j \neq i$, $2\omega^{ji} - \omega^{j-1,i} - \omega^{j+1,i} \leq 0$, and $\omega^{ij} > 0$.

$$
\begin{aligned}
& E(\sum_{i=1}^{n} \sum_{j=1}^{n} \omega^{ij} \epsilon_j \int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t)^2 \\
&= a_0^2 \sum_{i=1}^{n} \sum_{j=1}^{n} \{\omega^{ij}(2\omega^{ji} - \omega^{j-1,i} - \omega^{j+1,i}) E \int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt\} \\
&\leq a_0^2 B\Delta \sum_{i=1}^{n} \omega^{ii}(2\omega^{ii} - \omega^{i-1,i} - \omega^{i+1,i}) \\
&= O(n^{\frac{1}{2}})
\end{aligned}
$$

So all the conclusions claimed in the lemma hold.

5 Proof of Lemma 2

Let $\epsilon = (\epsilon_{\tau_1}, \epsilon_{\tau_2}, \ldots, \epsilon_{\tau_n})'$. Denote K^i, $K^{i,j}$, $K^{i,j,k}$, and $K^{i,j,k,l}$ the corresponding cumulants of $\{\epsilon_{\tau_i}\}$. Then $K^i = 0$, since the mean of ϵ_{τ_i} is zero. According to McCullagh (1987, Section 3.3) , we have

$$
\begin{aligned}
& var(\epsilon'\Omega\epsilon) \\
&= \sum_{i,j,k,l=1}^{n} \omega^{ij}\omega^{kl}\Big(K^{i,j,k,l} + 4K^i K^{j,k,l} + 2K^{i,k}K^{j,l} + 4K^i K^k K^{j,l}\Big) \\
&= \sum_{i,j,k,l=1}^{n} \omega^{ij}\omega^{kl}\Big(K^{i,j,k,l} + 2K^{i,k}K^{j,l}\Big) \\
&= \sum_{i,j,k,l=1}^{n} \omega^{ij}\omega^{kl}\Big(\text{cum}(\epsilon_{\tau_i}, \epsilon_{\tau_j}, \epsilon_{\tau_k}, \epsilon_{\tau_l}) + 2\text{cov}(\epsilon_{\tau_i}, \epsilon_{\tau_k})\text{cov}(\epsilon_{\tau_j}, \epsilon_{\tau_l})\Big) \\
&:= V_1(\omega, \omega) + V_2(\omega, \omega)
\end{aligned}
$$

where

$$
V_1(v, \omega) = \sum_{i,j,k,l=1}^{n} v^{ij}\omega^{kl}\text{cum}(\epsilon_{\tau_i}, \epsilon_{\tau_j}, \epsilon_{\tau_k}, \epsilon_{\tau_l})
$$

$$\begin{aligned}
V_2(v,\omega) &= \sum_{i,j,k,l=1}^{n} v^{ij}\omega^{kl} 2\mathrm{cov}(\epsilon_{\tau_i},\epsilon_{\tau_k})\mathrm{cov}(\epsilon_{\tau_j},\epsilon_{\tau_l}) \\
&= 2a_0^4 \sum_{i=1}^{N}\sum_{j=1}^{N}\{v^{ij}(\omega^{j-1,i-1}+\omega^{j-1,i+1}-2\omega^{j-1,i}+\omega^{j+1,i-1}+\omega^{j+1,i+1} \\
&\quad -2\omega^{j+1,i}-2(\omega^{j,i-1}+\omega^{j,i+1}-2\omega^{j,i}))\}
\end{aligned}$$

Recall that in Lemma 1 of Aït-Sahalia et al. (2005),

$$\begin{aligned}
&\mathrm{cum}(\epsilon_{\tau_i},\epsilon_{\tau_j},\epsilon_{\tau_k},\epsilon_{\tau_l}) \\
&= \begin{cases} 2\mathrm{cum}_4[U], & \text{if } i=j=k=l; \\ (-1)^{s(i,j,k,l)}\mathrm{cum}_4[U], & \text{if } \max(i,j,k,l)=\min(i,j,k,l)+1; \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}$$

where $s(i,j,k,l)$ denotes the number of indices among (i,j,k,l) that are equal to $\min(i,j,k,l)$.

So using this lemma, we can obtain

$$\begin{aligned}
&V_1(v,\omega) \\
&= \mathrm{cum}_4[U]\Big(2\sum_{i=1}^{n} v^{ii}\omega^{ii}+\sum_{i=1}^{n-1}(-4v^{i,i+1}\omega^{i+1,i+1}-4v^{i+1,i+1}\omega^{i,i+1} \\
&\quad +2v^{i,i}\omega^{i+1,i+1}+4v^{i,i+1}\omega^{i,i+1})\Big)
\end{aligned}$$

To facilitate calculation, we rewrite

$$\begin{aligned}
&V_2(\frac{\partial\omega}{\partial\sigma^2},\frac{\partial\omega}{\partial\sigma^2}) \\
&= (\frac{\partial\eta}{\partial\sigma^2})^2 V_2(\frac{\partial\omega}{\partial\eta},\frac{\partial\omega}{\partial\eta})+\frac{1}{\gamma^4}(\frac{\partial\gamma^2}{\partial\sigma^2})^2 V_2(\omega,\omega)-\frac{1}{\gamma^2}\frac{\partial\eta}{\partial\sigma^2}\frac{\partial\gamma^2}{\partial\sigma^2}\frac{\partial V_2(\omega,\omega)}{\partial\eta}
\end{aligned}$$

Then direct computation yields,

$$\begin{aligned}
V_2(\frac{\partial\omega}{\partial\eta},\frac{\partial\omega}{\partial\eta}) &= \frac{4a_0^4\{3(9-7n)\eta^2+3(19+3n)\eta^3+15(1+n)\eta^4-3(1+n)\eta^5\}}{3\gamma^4(-1-\eta)\eta^2(1-\eta)^6}+o(1) \\
&= \frac{a_0^4 n^{\frac{3}{2}}}{4a^3\sigma\sqrt{T}}+o(n^{\frac{3}{2}}) && (30) \\
V_2(\omega,\omega) &= \frac{4a_0^4}{\gamma^4(1-\eta)^4}\{-1-4\eta+\eta^2+n(1-\eta)(3-\eta)\}+o(1) \\
&= \frac{2na_0^4}{a^4}+o(n) && (31) \\
\frac{\partial V_2(\omega,\omega)}{\partial\eta} &= \frac{-8a_0^4\{4(-1+n)\eta-5(1+n)\eta^2+(1+n)\eta^3\}}{-\gamma^4\eta(1-\eta)^5}+o(1)
\end{aligned}$$

$$= \frac{5na_0^4}{2a^4} + o(n) \tag{32}$$

Note that

$$\begin{aligned}
\frac{\partial\eta}{\partial\sigma^2} &= \frac{1}{2}\sqrt{\frac{T}{a^2\sigma^2}}n^{-\frac{1}{2}} - \frac{T}{2a^2}n^{-1} + O(n^{-\frac{3}{2}}) \\
\frac{\partial\gamma^2}{\partial\sigma^2} &= \frac{1}{2}\sqrt{\frac{a^2T}{\sigma^2}}n^{-\frac{1}{2}} + \frac{T}{2}n^{-1} + O(n^{-\frac{3}{2}})
\end{aligned}$$

Combining (30), (31) and (32), we obtain,

$$V_2(\frac{\partial\omega}{\partial\sigma^2}, \frac{\partial\omega}{\partial\sigma^2}) = \frac{\sqrt{T}a_0^4 n^{\frac{1}{2}}}{16a^5\sigma^3} + o(n^{\frac{1}{2}}) \tag{33}$$

Similarly,

$$\begin{aligned}
\frac{\partial\eta}{\partial a^2} &= -\frac{\sqrt{a^2\sigma^2T}}{2a^4}n^{-\frac{1}{2}} + \frac{\sigma^2T}{2a^4}n^{-1} + O(n^{-\frac{3}{2}}) \\
\frac{\partial\gamma^2}{\partial a^2} &= 1 + \frac{\sqrt{a^2\sigma^2T}}{2a^2}n^{-\frac{1}{2}} + O(n^{-\frac{3}{2}})
\end{aligned}$$

Therefore,

$$\begin{aligned}
& V_2(\frac{\partial\omega}{\partial a^2}, \frac{\partial\omega}{\partial a^2}) \\
= \; & (\frac{\partial\eta}{\partial a^2})^2 V_2(\frac{\partial\omega}{\partial\eta}, \frac{\partial\omega}{\partial\eta}) + \frac{1}{\gamma^4}(\frac{\partial\gamma^2}{\partial a^2})^2 V_2(\omega,\omega) - \frac{1}{\gamma^2}\frac{\partial\eta}{\partial a^2}\frac{\partial\gamma^2}{\partial a^2}\frac{\partial V_2(\omega,\omega)}{\partial\eta} \\
= \; & \frac{2na_0^4}{a^8} + o(n)
\end{aligned}$$

It follows from the similar calculation as above (see also Aït-Sahalia et al., 2005, pp 396) that

$$\begin{aligned}
V_1(\frac{\partial\omega}{\partial\eta}, \frac{\partial\omega}{\partial\eta}) &= \frac{4n}{\gamma^4(1-\eta)^4}\mathrm{cum}_4[U] + o(n) = \frac{n}{4a^4}\mathrm{cum}_4[U] + o(n) \\
V_1(\omega,\omega) &= \frac{4n}{\gamma^4(1-\eta)^2}\mathrm{cum}_4[U] + o(n) = \frac{n}{a^4}\mathrm{cum}_4[U] + o(n) \\
\frac{\partial V_1(\omega,\omega)}{\partial\eta} &= \frac{8n}{\gamma^4(1-\eta)^3}\mathrm{cum}_4[U] + o(n) = \frac{n}{a^4}\mathrm{cum}_4[U] + o(n)
\end{aligned}$$

Therefore,

$$\begin{aligned}
& V_1(\frac{\partial\omega}{\partial\sigma^2}, \frac{\partial\omega}{\partial\sigma^2}) = O(1) \\
& V_1(\frac{\partial\omega}{\partial a^2}, \frac{\partial\omega}{\partial a^2}) = \frac{n\mathrm{cum}_4[U]}{a^8} + o(n)
\end{aligned}$$

Hence, combining the above equalities, we obtain

$$var(\epsilon'\frac{\partial\Omega}{\partial\sigma^2}\epsilon) = V_1(\frac{\partial\omega}{\partial\sigma^2},\frac{\partial\omega}{\partial\sigma^2}) + V_2(\frac{\partial\omega}{\partial\sigma^2},\frac{\partial\omega}{\partial\sigma^2}) = \frac{\sqrt{T}a_0^4n^{\frac{1}{2}}}{16a^5\sigma^3} + o(n^{\frac{1}{2}})$$
$$var(\epsilon'\frac{\partial\Omega}{\partial a^2}\epsilon) = V_1(\frac{\partial\omega}{\partial a^2},\frac{\partial\omega}{\partial a^2}) + V_2(\frac{\partial\omega}{\partial a^2},\frac{\partial\omega}{\partial a^2}) = \frac{n(2a_0^4+\mathrm{cum}_4[U])}{a^8} + o(n)$$

This concludes the proof.

6 Proof of Theorem 4

We want to prove it by verifying the conditions of Theorem 2. It follows from Lemmas 1 and 2 that

$$\begin{aligned}
\Psi_n^1 - \bar{\Psi}_n^1 &= \frac{1}{2\sqrt{n}}\{O_p(1) + O_p(n^{\frac{1}{4}}) + O_p(n^{\frac{1}{4}}) + O_p(n^{\frac{1}{4}})\} = O_p(n^{-\frac{1}{4}}) \\
\Psi_n^2 - \bar{\Psi}_n^2 &= \frac{1}{2n}\{O_p(1) + O_p(n^{\frac{1}{4}}) + O_p(n^{\frac{1}{4}}) + O_p(n^{\frac{1}{2}})\} = O_p(n^{-\frac{1}{2}})
\end{aligned}$$

So far, we have shown that $|\Psi_n^i - \bar{\Psi}_n^i| \xrightarrow{\mathrm{P}} 0$, for any $\theta \in \Theta$, $i = 1, 2$. To prove uniform convergence in probability, we need stochastic equicontinuity of $\Psi_n^i - \bar{\Psi}_n^i$. As argued in Newey and McFadden (1994, pp 2133-2134) (see also Rockafellar (1970, Theorem 10.8)), the pointwise convergence of concave functions implies uniform convergence. Note that $\Psi_n^i - \bar{\Psi}_n^i$ can be regarded as the difference of two concave functions from (12), then a slightly modified proof of Theorem 10.8 in Rockafellar (1970), using triangle inequalities, still gives rise to uniform convergence. Therefore, as in Andersen and Gill (1982), we arrive at

$$\sup_{\theta\in\Theta}\|\Psi_n(\theta) - \bar{\Psi}_n(\theta)\| \xrightarrow{\mathrm{P}} 0.$$

Next, we show the identifiability condition holds.

$$\begin{aligned}
\bar{\Psi}_n^2 &= \frac{1}{2n}\{tr(\Omega^{-1}\frac{\partial\Omega}{\partial a^2}) + tr(\frac{\partial\Omega^{-1}}{\partial a^2}\Sigma_0)\} \\
&= \frac{1}{n}\{(tr\Omega^{-1} - tr\Omega^{-1}J) + a_0^2(tr\frac{\partial\Omega^{-1}}{\partial a^2} - tr\frac{\partial\Omega^{-1}}{\partial a^2}J) \\
&\quad + \frac{1}{2}\sum_{i=1}^{n}\frac{\partial\omega^{ii}}{\partial a^2}\int_{\tau_{i-1}}^{\tau_i}\sigma_t^2 dt\}
\end{aligned}$$

where $J = (J_{ij})$ where $J_{i-1,i} = 1$, and the other components of J is 0.

$$
\begin{aligned}
tr\Omega^{-1} - tr(\Omega^{-1}J) &= \frac{n(1+\eta)(1+\eta^{2n+1}) - \eta(1+\eta^{2n}) - \frac{2\eta^2(1-\eta^{2n-1})}{1-\eta}}{\gamma^2(1-\eta^2)(1-\eta^{2n+2})} \\
&= \frac{n}{2a^2} - \frac{\sqrt{Ta^2\sigma^2}}{4a^4} n^{\frac{1}{2}} + O(1) \qquad (34)
\end{aligned}
$$

So

$$
tr\frac{\partial\Omega^{-1}}{\partial a^2} - tr(\frac{\partial\Omega^{-1}}{\partial a^2}J) = -\frac{n}{2a^4} + \frac{3\sqrt{a^2\sigma^2T}}{8a^6} n^{\frac{1}{2}} + O(1) \qquad (35)
$$

On the other hand, choose $K = n^{\frac{2}{3}}$, and let [7] $\omega^m = \omega^{K,K}$ and $\omega^M = \omega^{\frac{n+1}{2},\frac{n+1}{2}}$. By (28), we have

$$
\begin{aligned}
1 \le \frac{\omega^M}{\omega^m} &= \frac{(1-\eta^{n+1})(1-\eta^{n+1})}{(1-\eta^{2(n^{\frac{2}{3}})})(1-\eta^{2n-2(n^{\frac{2}{3}})+2})} \\
&= \frac{(1-e^{-\sqrt{\frac{\sigma^2T}{a^2}}n^{\frac{1}{2}}+O(n^{-\frac{1}{2}})})(1-e^{-\sqrt{\frac{\sigma^2T}{a^2}}n^{\frac{1}{2}}+O(n^{-\frac{1}{2}})})}{(1-e^{-\sqrt{\frac{\sigma^2T}{a^2}}n^{\frac{1}{6}}+O(n^{-\frac{5}{6}})})(1-e^{-2\sqrt{\frac{\sigma^2T}{a^2}}n^{\frac{1}{2}}+O(n^{-\frac{1}{2}})})} \\
&\to 1
\end{aligned}
$$

Therefore, for any $K \le i \le n-K$, $\omega^{ii} = \omega^m(1+o(1))$; for $i < K$ or $i > n-K$, ω^{ii} is donimated by ω^m, and the integration is over an interval which shrinks at the rate $K/n = n^{-1/3}$, so

$$
\sum_{i=1}^{n} \omega^{ii} \int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt = \omega^m \int_0^T \sigma_t^2 dt(1+o_p(1)) = \frac{n^{\frac{1}{2}}(a^2)^{-\frac{1}{2}}}{2\sqrt{\sigma^2T}} \int_0^T \sigma_t^2 dt(1+o_p(1))
$$

Similarly, for $K \le i \le n-K$,

$$
\begin{aligned}
\frac{\partial\omega^{ii}}{\partial a^2} &= \frac{\partial\omega^{ii}}{\partial\eta}\frac{\partial\eta}{\partial a^2} - \frac{\omega^{ii}}{\gamma^2}\frac{\partial\gamma^2}{\partial a^2} = (\frac{2\eta}{1-\eta^2}\frac{\partial\eta}{\partial a^2} - \frac{1}{\gamma^2}\frac{\partial\gamma^2}{\partial a^2})\omega^{ii}(1+o(1)) \\
&= -\frac{1}{2a^2}\omega^{ii}(1+o(1))
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial\omega^{ii}}{\partial\sigma^2} &= \frac{\partial\omega^{ii}}{\partial\eta}\frac{\partial\eta}{\partial\sigma^2} - \frac{\omega^{ii}}{\gamma^2}\frac{\partial\gamma^2}{\partial\sigma^2} = (\frac{2\eta}{1-\eta^2}\frac{\partial\eta}{\partial\sigma^2} - \frac{1}{\gamma^2}\frac{\partial\gamma^2}{\partial\sigma^2})\omega^{ii}(1+o(1)) \\
&= -\frac{1}{2\sigma^2}\omega^{ii}(1+o(1))
\end{aligned}
$$

[7] Here and subsequently, $n^{\frac{2}{3}}$ can be replaced by $n^{\frac{1}{2}+\delta}$ with any $0 < \delta < \frac{1}{2}$.

while for $i < K$ or $i > n-K$, $\frac{\partial \omega^{ii}}{\partial a^2}$ and $\frac{\partial \omega^{ii}}{\partial \sigma^2}$ are dominated by ω^{ii}. Therefore,

$$\sum_{i=1}^{n} \frac{\partial \omega^{ii}}{\partial a^2} \int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt \quad = \quad -\frac{n^{\frac{1}{2}}(a^2)^{-\frac{3}{2}}}{4\sqrt{\sigma^2 T}} \int_0^T \sigma_t^2 dt (1 + o_p(1)) \tag{36}$$

and

$$\sum_{i=1}^{n} \frac{\partial \omega^{ii}}{\partial \sigma^2} \int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt \quad = \quad -\frac{n^{\frac{1}{2}}(\sigma^2)^{-\frac{3}{2}}}{4\sqrt{a^2 T}} \int_0^T \sigma_t^2 dt (1 + o_p(1)) \tag{37}$$

By calculation, we can also obtain

$$\bar{\Psi}_n^2 = \left(\frac{1}{2a^2} - \frac{a_0^2}{2a^4}\right) + \left(\frac{3a_0^2\sqrt{\sigma^2 T}}{8a^5} - \frac{\sqrt{\sigma^2 T}}{4a^3} - \frac{\int_0^T \sigma_t^2 dt}{8a^3\sqrt{\sigma^2 T}}\right) n^{-\frac{1}{2}} + o_p(n^{-\frac{1}{2}}) \tag{38}$$

Hence, if we solve $\bar{\Psi}_n^2 = 0$, we get

$$a_n^{2*} \quad = \quad a_0^2 + \left(\frac{3a_0^2\sqrt{\sigma_n^{2*}T}}{4a_n^*} - \frac{\sqrt{\sigma_n^{2*}T}a_n^*}{2} - \frac{a_n^* \int_0^T \sigma_t^2 dt}{4\sqrt{\sigma_n^{2*}T}}\right) n^{-\frac{1}{2}} + o_p(n^{-\frac{1}{2}}) \tag{39}$$

We assume all these likelihood estimates are bounded almost surely, since the parameter space is itself bounded. Also,

$$\begin{aligned}
\frac{\partial \bar{\Psi}_n^2}{\partial \sigma^2} &= \left(\frac{\int_0^T \sigma_t^2 dt}{16a^3\sqrt{T\sigma^6}} + \frac{3a_0^2}{16a^5\sqrt{\sigma^2 T}} - \frac{T}{8a^3\sqrt{\sigma^2 T}}\right) n^{-\frac{1}{2}} + o_p(n^{-\frac{1}{2}}) \\
\frac{\partial \bar{\Psi}_n^2}{\partial a^2} &= \frac{a_0^2}{a^6} - \frac{1}{2a^4} + \left(\frac{3\int_0^T \sigma_t^2 dt}{16a^5\sqrt{\sigma^2 T}} - \frac{15a_0^2\sqrt{\sigma^2 T}}{16a^7} + \frac{3\sqrt{\sigma^2 T}}{8a^5}\right) n^{-\frac{1}{2}} + o_p(n^{-\frac{1}{2}})
\end{aligned}$$

On the other hand, let $\Gamma := \Sigma - \Omega$ be the diagram matrix with the i^{th} element of the diagonal $\Gamma_i := \int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt - \sigma^2\Delta$, then

$$\begin{aligned}
&\bar{\Psi}_n^1(\sigma^2, a^2) = \frac{1}{2\sqrt{n}}\left\{tr\left(\Omega^{-1}\frac{\partial \Omega}{\partial \sigma^2}\right) + \frac{\partial tr(\Omega^{-1}\Sigma_0)}{\partial \sigma^2}\right\} \\
= \; & \frac{1}{2\sqrt{n}}\left\{tr\left(\Omega^{-1}\frac{\partial \Omega}{\partial \sigma^2}\right) + \frac{\partial tr(\Omega^{-1}(\Omega + (2I - J - J')(a^2 - a_0^2) + \Gamma))}{\partial \sigma^2}\right\} \\
= \; & \frac{1}{2\sqrt{n}}\left\{tr\left(\Omega^{-1}\frac{\partial \Omega}{\partial \sigma^2}\right) + \sum_{i=1}^{n} \frac{\partial(\omega^{ii}\Gamma_i)}{\partial \sigma^2}\right\} - \frac{(a^2 - a_0^2)T}{2a^2 n^{\frac{3}{2}}} \sum_{i=1}^{n}\left(\omega^{ii} + \sigma^2\frac{\partial \omega^{ii}}{\partial \sigma^2}\right) \\
= \; & \frac{1}{2\sqrt{n}}\left\{\Delta\sum_{i=1}^{n}\omega^{ii} + \sum_{i=1}^{n}\frac{\partial \omega^{ii}}{\partial \sigma^2}\Gamma_i - \Delta\sum_{i=1}^{n}\omega^{ii}\right\} - \frac{\sqrt{T}}{8a^3\sigma}(a^2 - a_0^2) + O_p(n^{-\frac{1}{2}}) \\
= \; & \frac{1}{2\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \omega^{ii}}{\partial \sigma^2}\Gamma_i - \frac{\sqrt{T}}{8a^3\sigma}(a^2 - a_0^2) + O_p(n^{-\frac{1}{2}})
\end{aligned}$$

$$= \quad -\frac{1}{8a\sigma^3\sqrt{T}}(\int_0^T \sigma_t^2 dt - \sigma^2 T)(1+o(1)) - \frac{\sqrt{T}}{8a^3\sigma}(a^2 - a_0^2) + O_p(n^{-\frac{1}{2}}) \tag{40}$$

where the last equality is given by (37).

Set $\bar{\Psi}_n^1(\sigma^2, a^2) = 0$, that is,

$$\int_0^T \sigma_t^2 dt - \sigma_n^{2*} T \quad = \quad -\frac{\sigma_n^{2*} T}{a_n^{2*}}(a_n^{2*} - a_0^2) + O_p(n^{-\frac{1}{2}}) = O_p(n^{-\frac{1}{2}}) \tag{41}$$

Again, it follows from direct calculation that

$$\begin{aligned}
\frac{\partial \bar{\Psi}_n^1}{\partial \sigma^2} &= \frac{(-a_0^2 + a^2)\sqrt{T}}{16a^3\sigma^3} + \frac{\sqrt{T}}{8a\sigma^3} + \frac{3(\int_0^T \sigma_t^2 dt - T\sigma^2)}{16a\sigma^5\sqrt{T}} + o_p(1) \\
\frac{\partial \bar{\Psi}_n^1}{\partial a^2} &= \frac{3(-a_0^2 + a^2)\sqrt{T}}{16a^5\sigma} - \frac{\sqrt{T}}{8a^3\sigma} + \frac{\int_0^T \sigma_t^2 dt - T\sigma^2}{16a^3\sigma^3\sqrt{T}} + o_p(1)
\end{aligned}$$

As $n \to \infty$, $\theta_n^* \xrightarrow{\text{P}} \theta_0$, so with probability approaching 1,

$$\min_{\theta \in \Theta: \|\theta - \theta_n^*\| \geq \epsilon} \|\bar{\Psi}_n(\theta)\| \geq \min_{\theta \in \Theta: \|\theta - \theta_0\| \geq \frac{\epsilon}{2}} \|\bar{\Psi}_n(\theta)\|$$

Observe that,

$$\begin{aligned}
& \|\bar{\Psi}_n(\theta)\| = |\bar{\Psi}_n^1|^2 + |\bar{\Psi}_n^2|^2 \\
= \quad & \frac{1}{4a^4}\{1 - \frac{a_0^2}{a^2} + (\frac{3a_0^2\sqrt{\sigma^2 T}}{8a^5} - \frac{\sqrt{\sigma^2 T}}{4a^3} - \frac{\int_0^T \sigma_t^2 dt}{8a^3\sqrt{\sigma^2 T}})n^{-\frac{1}{2}} + o(n^{-\frac{1}{2}})\}^2 \\
& +\{\frac{1}{8a\sigma^3\sqrt{T}}(\int_0^T \sigma_t^2 dt - \sigma^2 T)(1+o(1)) + \frac{\sqrt{T}}{8a^3\sigma}(a^2 - a_0^2) + O(n^{-\frac{1}{2}})\}^2 \\
= \quad & \frac{1}{4a^4}(1 - \frac{a_0^2}{a^2})^2 + \frac{T}{64a^2\sigma^2}\{\frac{1}{\sigma^2 T}(\int_0^T \sigma_t^2 dt - \sigma^2 T) + (1 - \frac{a_0^2}{a^2})\}^2 + o_p(1) \\
> \quad & \epsilon_0 + o_p(1)
\end{aligned}$$

when $\|\theta - \theta_0\| = \sqrt{(\sigma^2 - \frac{1}{T}\int_0^T \sigma_t^2 dt)^2 + (a^2 - a_0^2)^2} > \frac{\epsilon}{2}$ and n goes to ∞. Hence,

$$\begin{aligned}
P(\min_{\theta \in \Theta: \|\theta - \theta_n^*\| \geq \epsilon} \|\bar{\Psi}_n(\theta)\| > \frac{\epsilon_0}{2}) \quad &> \quad P(\min_{\theta \in \Theta: \|\theta - \theta_0\| \geq \frac{\epsilon}{2}} \|\bar{\Psi}_n(\theta)\| > \frac{\epsilon_0}{2}) \\
&> \quad P(-o_p(1) < \frac{\epsilon_0}{2}) \to 1
\end{aligned}$$

Therefore, it follows that the identifiability condition holds, hence by Theorem 2, $\hat{\sigma}^2 - \sigma_n^{2*} \xrightarrow{\text{P}} 0$, and $\hat{a}^2 - a_n^{2*} \xrightarrow{\text{P}} 0$. This concludes the proof by (39) and (41).

7 Proof of Lemma 3

The argument is very similar to the proof of Theorem 1 given in the appendix of Barndorff-Nielsen et al. (2008), which involves the concept of stable convergence in law. The details about it have been discussed in Jacod and Shiryaev (2003) and Jacod (2007).

Assume for convenience that $\omega^{ij} = 0$, for $i, j < 1$ or $i, j > n$. Also, denote

$$\begin{aligned}
M_1^{(\beta)} &= \sum_{i=1}^{n} \frac{\partial \omega^{ii}}{\partial \beta} \{(\Delta_i^n X)^2 - \int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt\} \\
M_2^{(\beta)} &= \sum_{i=1}^{n} \sum_{j<i} \frac{\partial \omega^{ij}}{\partial \beta} \Delta_i^n X \Delta_j^n X \\
M_3^{(\beta)} &= \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial \omega^{ij}}{\partial \beta} \epsilon_j \Delta_i^n X := \sum_{i=1}^{n} \iota_i^{(\beta)} \Delta_i^n X \\
M_4^{(\beta)} &= \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial \omega^{ij}}{\partial \beta} (\epsilon_i \epsilon_j - E\epsilon_i \epsilon_j)
\end{aligned}$$

where $\Delta_i^n X = X_{\tau_i} - X_{\tau_{i-1}}$ and $\iota_i^{(\beta)} = \sum_{j=1}^{n} \frac{\partial \omega^{ij}}{\partial \beta} \epsilon_j$.

We begin with $M_2^{(\sigma^2)}$. Pick $K = n^{\frac{2}{3}}$, and consider $K \leq i \leq n - K$. When $|i-j| = O(n^{\frac{1}{2}+\delta})$, $\omega^{ij} \to 0$ exponentially, for any $0 < \delta < \frac{1}{2}$. Rewrite it as:

$$n^{-\frac{3}{4}} M_2^{(\sigma^2)} \approx \Delta \sum_i f(\frac{\Delta_i^n X}{\sqrt{\Delta}}, \frac{\Delta_{i-1}^n X}{\sqrt{\Delta}}, ..., \frac{\Delta_{i-K}^n X}{\sqrt{\Delta}})$$

where

$$f(x_i, x_{i-1}, ..., x_{i-K}) = n^{-\frac{3}{4}} \sum_{h=1}^{K} \frac{\partial \omega^{i,i-h}}{\partial \sigma^2} x_i x_{i-h}$$

It follows from Theorem 7.1 in Jacod (2007) that, $\Delta^{-\frac{1}{2}} \cdot n^{-\frac{3}{4}} M_2^{(\sigma^2)}$ converges stably in law, and its asymptotic variance can be calculated using formula (7.2) and (7.3) in Jacod (2007):

$$Avar = \lim_{n \to \infty} \sum_i \sigma_{\tau_i}^4 R_{i,n}(f) \Delta$$

where

$$\begin{aligned}
R_{i,n}(f) &= n^{-\frac{3}{2}} \{E(\sum_{h=1}^{K} \frac{\partial \omega^{i,i-h}}{\partial \sigma^2} U_i U_{i-h})^2 - (E \sum_{h=1}^{K} \frac{\partial \omega^{i,i-h}}{\partial \sigma^2} U_i U_{i-h})^2\} \\
&= n^{-\frac{3}{2}} \sum_{i-K \leq j < i} (\frac{\partial \omega^{ij}}{\partial \sigma^2})^2
\end{aligned}$$

and U_is are i.i.d. standard Gaussian random variables.

For any $K \leq i \leq n-K$, direct calculation yields

$$n^{-\frac{3}{2}} \sum_{i-K \leq j < i} (\frac{\partial \omega^{ij}}{\partial \sigma^2})^2 \longrightarrow \frac{5}{64 T^{\frac{3}{2}} \sigma^7 a}$$

By the same argument used in the proof of Theorem 4, we have

$$n^{-\frac{1}{4}} M_2^{(\sigma^2)} \xrightarrow{\mathcal{L}_X} MN(0, \frac{5}{64\sqrt{T}\sigma^7 a} \int_0^T \sigma_t^4 dt) \tag{42}$$

Since we have shown in Lemma 1 that $n^{-\frac{1}{4}} M_1^{(\sigma^2)} \xrightarrow{\mathrm{P}} 0$, it follows from Lemma 2 in Barndorff-Nielsen et al. (2008) that

$$n^{-\frac{1}{4}} (M_1^{(\sigma^2)} + 2M_2^{(\sigma^2)}) \xrightarrow{\mathcal{L}_X} MN(0, \frac{5}{16\sqrt{T}\sigma^7 a} \int_0^T \sigma_t^4 dt) \tag{43}$$

As to $M_3^{(\sigma^2)}$, we at first notice that

$$E((\iota_i^{(\sigma)})^2 | \sigma(X)) = a_0^2 \sum_{j=1}^{n} \frac{\partial \omega^{ij}}{\partial \sigma^2} (2\frac{\partial \omega^{ij}}{\partial \sigma^2} - \frac{\partial \omega^{i,j-1}}{\partial \sigma^2} - \frac{\partial \omega^{i,j+1}}{\partial \sigma^2})$$

and that for $K \leq i \leq n-K$,

$$n^{-\frac{1}{2}} \sum_{j=1}^{n} \frac{\partial \omega^{ij}}{\partial \sigma^2} (2\frac{\partial \omega^{ij}}{\partial \sigma^2} - \frac{\partial \omega^{i,j-1}}{\partial \sigma^2} - \frac{\partial \omega^{i,j+1}}{\partial \sigma^2}) \longrightarrow \frac{1}{8\sqrt{T}\sigma^5 a^3}$$

and that $\{(\sum_{i=1}^{n} (\Delta_i^n X)^2)^{-\frac{1}{2}} \Delta_i^n X\}$ become uniformly asymptotically negligible. Therefore, by the standard central limit theorem, conditional on the filtration $\sigma(X)$, we have

$$(\sum_{i=1}^{n} (\Delta_i^n X)^2)^{-\frac{1}{2}} n^{-\frac{1}{4}} M_3^{(\sigma^2)} \xrightarrow{\mathcal{L}} N(0, \frac{a_0^2}{8\sqrt{T}\sigma^5 a^3})$$

Since $\sum_{i=1}^{n} (\Delta_i^n X)^2 \xrightarrow{\mathrm{P}} \int_0^T \sigma_t^2 dt$, and by Lemma 1 and Proposition 5 in Barndorff-Nielsen et al. (2008), we can conclude that

$$n^{-\frac{1}{4}} M_3^{(\sigma^2)} \xrightarrow{\mathcal{L}_X} N(0, \frac{a_0^2}{8\sqrt{T}\sigma^5 a^3} \int_0^T \sigma_t^2 dt) \tag{44}$$

and $n^{-\frac{1}{4}} (M_1^{(\sigma^2)} + 2M_2^{(\sigma^2)})$ and $n^{-\frac{1}{4}} M_3^{(\sigma^2)}$ jointly converge $\sigma(X)$-stably in law.

Also, as to the noise part, conditional on the filtration $\sigma(X)$,

$$n^{-\frac{1}{4}} M_4^{(\sigma^2)} \xrightarrow{\mathcal{L}} N(0, \frac{\sqrt{T} a_0^4}{16 a^5 \sigma^3}) \tag{45}$$

The same reasoning as above yields

$$
\begin{aligned}
n^{\frac{1}{4}}(\Psi_n^1 - \bar{\Psi}_n^1) &= n^{-\frac{1}{4}}(M_1^{(\sigma^2)} + 2M_2^{(\sigma^2)} + M_3^{(\sigma^2)} + M_4^{(\sigma^2)}) \\
&\xrightarrow{\mathcal{L}_X} MN(0, \frac{5}{16\sqrt{T}\sigma^7 a}\int_0^T \sigma_t^4 dt + \frac{a_0^2}{8\sqrt{T}\sigma^5 a^3}\int_0^T \sigma_t^2 dt + \frac{\sqrt{T}a_0^4}{16a^5\sigma^3})
\end{aligned} \tag{46}
$$

As to $\Psi_n^2 - \bar{\Psi}_n^2$, since the noise part dominates, it implies that

$$
n^{\frac{1}{2}}(\Psi_n^2 - \bar{\Psi}_n^2) \xrightarrow{\mathcal{L}} MN(0, \frac{2a_0^4 + \mathrm{cum}_4[U]}{4a^8}) \tag{47}
$$

Last, the covariance of the $\Psi_n^1 - \bar{\Psi}_n^1$ and $\Psi_n^2 - \bar{\Psi}_n^2$ is of the order $O(n^{-1})$, thus the joint central limit theorem follows from Proposition 5 in Barndorff-Nielsen et al. (2008).

8 Proof of Theorem 5

Now we derive the central limit theorem of the estimators. Because

$$
\Psi_n^i(\omega, \hat{\theta}) - \Psi_n^i(\omega, \theta_n^*) = \nabla\Psi_n^i(\omega, \tilde{\theta}_i)(\hat{\theta} - \theta_n^*)
$$

where $\tilde{\theta}_i$ is between $\hat{\theta}$ and θ_n^*, and by Lemma 1 in Barndorff-Nielsen et al. (2008), it is clear that we only need to verify the assumptions of Theorem 3. In fact, because of (12), we will consider $Y'\Omega Y$ and $tr(\Omega\Sigma_0)$ and their derivatives multiplied by proper normalizations. It is straightforward to check that these functions are either convex or concave, which guarantees stochastic equicontinuity (see the proof of Theorem 4).

Combining Theorem 4, we have

$$
\begin{aligned}
&\begin{pmatrix} n^{\frac{1}{4}}(\hat{\sigma}^2 - \sigma_n^{2*}) \\ n^{\frac{1}{2}}(\hat{a}^2 - a_n^{2*}) \end{pmatrix} \\
= &\begin{pmatrix} n^{\frac{1}{4}} & 0 \\ 0 & n^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} \frac{\sqrt{T}}{8a_n^*\sigma_n^{3*}} & -\frac{\sqrt{T}}{8a^{3*}\sigma^*} \\ 0 & \frac{a_0^2}{a^{6*}} - \frac{1}{2a^{4*}} \end{pmatrix}^{-1} \begin{pmatrix} \Psi_n^1(\omega, \theta_n^*) - \bar{\Psi}_n^1(\omega, \theta_n^*) \\ \Psi_n^2(\omega, \theta_n^*) - \bar{\Psi}_n^2(\omega, \theta_n^*) \end{pmatrix} + \begin{pmatrix} o_p(1) \\ o_p(1) \end{pmatrix} \\
\xrightarrow{\mathcal{L}_X} \; &MN\Bigg(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{5a_0\int_0^T \sigma_t^4 dt}{T(\int_0^T \sigma_t^2 dt)^{\frac{1}{2}}} + \frac{3a_0(\int_0^T \sigma_t^2 dt)^{\frac{3}{2}}}{T^2} & 0 \\ 0 & 2a_0^4 + \mathrm{cum}_4[U] \end{pmatrix}\Bigg)
\end{aligned}
$$

Note from (39) and (41) that because of the consistency results above, $a_n^{2*} - a_0^2 = o_p(n^{-\frac{1}{2}})$, $\sigma_n^{2*} - \frac{1}{T}\int_0^T \sigma_t^2 dt = O_p(n^{-\frac{1}{2}})$, therefore, the potential asymptotic biases are eliminated, which concludes the proof.

9 Proof of Theorem 6

From the likelihood function (2), we can obtain the score functions:

$$tr(\Omega^{-1}\Lambda) - tr(Y'\Omega^{-1}\Lambda\Omega^{-1}Y) = 0$$
$$tr(\Omega^{-1}) - tr(Y'\Omega^{-2}Y) = 0$$

where $a^2\Lambda = \Omega - \sigma^2\Delta I$.

On the other hand, the following equalities hold:

$$\begin{aligned}
a^2 tr(\Omega^{-1}\Lambda) &= tr(\Omega^{-1}(\Omega - \sigma^2\Delta I)) = n - \sigma^2\Delta tr(\Omega^{-1}) \\
a^2 tr(\Omega^{-2}\Lambda) &= tr(\Omega^{-2}(\Omega - \sigma^2\Delta I)) = tr(\Omega^{-1}) - \sigma^2\Delta tr(\Omega^{-2}) \\
a^4 tr(\Omega^{-2}\Lambda^2) &= n - 2\sigma^2\Delta tr(\Omega^{-1}) + \sigma^4\Delta^2 tr(\Omega^{-2})
\end{aligned}$$

As a result,

$$\begin{aligned}
\sigma^2\Delta &= \frac{tr(\Omega^{-2}\Lambda)tr(\Lambda\Omega^{-1}) - tr(\Omega^{-2}\Lambda^2)tr(\Omega^{-1})}{(tr(\Omega^{-2}\Lambda))^2 - tr(\Omega^{-2})\cdot tr(\Omega^{-2}\Lambda^2)} \\
a^2 &= \frac{tr(\Omega^{-2}\Lambda)tr(\Omega^{-1}) - tr(\Omega^{-2})tr(\Omega^{-1}\Lambda)}{(tr(\Omega^{-2}\Lambda))^2 - tr(\Omega^{-2})\cdot tr(\Omega^{-2}\Lambda^2)}
\end{aligned}$$

Plugging in the score functions gives the representation. Clearly, the two quadratic forms are not co-linear. The second claim is obvious. In fact, we can write $\Omega = \sigma^2 T\tilde{\Omega}$, and $\tilde{\Omega}$ only depends on λ. Plugging it into the representation is amount to replacing Ω with $\tilde{\Omega}$ directly.

10 Proof of Theorem 7

According to (28), we have that for $K \le i, j \le n - K$,

$$(\Omega^{-1})_{i,j} = \omega^{ij} \approx \frac{(-\eta)^{|i-j|}}{\gamma^2(1-\eta^2)}$$

with the difference exponentially small. Therefore, we can further deduce that

$$\begin{aligned}
(\Omega^{-2})_{i,j} &= \sum_l \omega^{il}\omega^{lj} \approx \frac{(-\eta)^{|i-j|}}{\gamma^4(1-\eta^2)^2}(\frac{2\eta^2}{1-\eta^2} + |i-j|) \\
&\approx \frac{n(-\eta)^{|i-j|}}{4a^4}\frac{a^2}{\sigma^2 T}\Big(\frac{a}{\sigma T^{\frac{1}{2}}}n^{\frac{1}{2}} + |i-j|\Big) \\
(\Omega^{-1}\Lambda\Omega^{-1})_{i,j} &= \frac{1}{a^2}\Big((\Omega^{-1})_{i,j} - \sigma^2\Delta(\Omega^{-2})_{i,j}\Big)
\end{aligned}$$

$$\approx \quad \frac{(-\eta)^{|i-j|}}{4a^4}\Big(\frac{a}{\sigma T^{\frac{1}{2}}}n^{\frac{1}{2}} - |i-j|\Big)$$

On the other hand, direct calculations deduce that

$$\begin{aligned}
tr(\Omega^{-2}) &= \frac{n^{\frac{5}{2}}}{4a\sigma^3 T^{\frac{3}{2}}} + O(n^2) \\
tr(\Omega^{-2}\Lambda) &= \frac{n^{\frac{3}{2}}}{4a^3\sigma\sqrt{T}} + O(n) \\
tr(\Omega^{-2}\Lambda^2) &= \frac{n}{a^4} + O(n^{\frac{1}{2}})
\end{aligned}$$

Therefore, it follows from the representation that

$$W_{1,i,j} \approx (1 + \frac{|i-j|}{\lambda \cdot n^{\frac{1}{2}}})(-\eta)^{|i-j|} \approx (1 + \frac{|i-j|}{\lambda \cdot n^{\frac{1}{2}}})e^{-\lambda^{-1}\cdot n^{-\frac{1}{2}}|i-j|}$$

11 Tables and Figures

Table 1: In this table, we report the finite sample quartiles of the infeasible standardized QMLE, which employs the theoretic asymptotic variance. The benchmark quartiles are those for the limit distribution $N(0,1)$.

No. obs	Mean	Stdv.	0.5%	2.5%	5%	95%	97.5%	99.5%
σ^2								
130	0.0099	1.1867	0.0001	0.0112	0.0422	0.9073	0.9339	0.9672
390	-0.0164	1.0807	0.0010	0.0170	0.0455	0.9273	0.9541	0.9830
780	0.0045	1.0861	0.0015	0.0175	0.0469	0.9248	0.9539	0.9818
1170	0.0044	1.0502	0.0022	0.0175	0.0454	0.9311	0.9579	0.9877
2340	-0.0017	1.0328	0.0025	0.0183	0.0440	0.9353	0.9639	0.9890
4680	-0.0032	1.0112	0.0025	0.0167	0.0432	0.9403	0.9660	0.9898
23400	-0.0057	1.0131	0.0031	0.0236	0.0492	0.9464	0.9706	0.9912
a^2								
130	-0.0084	1.2284	0.0135	0.0464	0.0818	0.9079	0.9400	0.9779
390	-0.0069	1.1146	0.0079	0.0355	0.0641	0.9236	0.9578	0.9859
780	0.0015	1.0909	0.0073	0.0334	0.0631	0.9315	0.9609	0.9878
1170	0.0037	1.0656	0.0057	0.0297	0.0558	0.9356	0.9627	0.9893
2340	-0.0186	1.0350	0.0066	0.0294	0.0554	0.9439	0.9707	0.9934
4680	-0.0063	1.0319	0.0063	0.0290	0.0551	0.9440	0.9700	0.9939
23400	-0.0026	1.0163	0.0047	0.0279	0.0533	0.9465	0.9712	0.9932

Table 2: This table reports the estimates for $100 \cdot (\hat{\sigma}^2 - \frac{1}{T}\int_0^T \sigma_t^2 dt)$, where $\hat{\sigma}^2$ is given by the QMLE and various implementations of the Tukey-Hanning$_2$ kernel respectively. Among them, RK_1 and RK_2 have the theoretical bandwidth, while RK_1 and RK_3 employ out-of-period data.

		1 sec	5 sec	10 sec	20 sec	30 sec	1 min	3 min
	Bias	0.0189	0.0480	-0.0040	-0.0065	-0.0341	-0.0452	-0.0838
QMLE	Stdv	1.1860	1.8172	2.1611	2.6040	2.9270	3.5828	4.9203
	RMSE	1.1861	1.8178	2.1611	2.6040	2.9272	3.5831	4.9210
	Bias	0.0178	0.0465	0.0074	0.0092	-0.0079	-0.0192	0.0003
RK_1	Stdv	1.2328	1.8844	2.2424	2.6835	3.0087	3.6851	5.0185
	RMSE	1.2329	1.8850	2.2424	2.6836	3.0087	3.6851	5.0185
	Bias	-0.1620	-0.3659	-0.5794	-0.8407	-1.0288	-1.4913	-2.5874
RK_2	Stdv	1.2197	1.8468	2.1847	2.5873	2.8704	3.4239	4.4245
	RMSE	1.2305	1.8827	2.2602	2.7205	3.0492	3.7345	5.1255
	Bias	0.0186	0.0247	-0.0048	0.0699	-0.0452	0.0009	0.3615
RK_3	Stdv	1.8603	2.7555	3.2438	3.7953	4.1695	4.8708	6.2255
	RMSE	1.8604	2.7556	3.2438	3.7960	4.1697	4.8708	6.2360
	Bias	-0.0641	-0.1568	-0.2591	-0.3008	-0.4934	-0.6316	-0.7731
RK_4	Stdv	1.8546	2.7293	3.2091	3.7537	4.0768	4.7444	5.9674
	RMSE	1.8557	2.7337	3.2196	3.7658	4.1066	4.7863	6.0173

Table 3: Summary Statistics of the Log Returns of the Euro/US Dollar Future

Avg No of Obs	Avg Freq	Mean	Std Err	1st Lag	2nd Lag
19657	4.44s	6.48e-09	6.12e-05	-0.075	0.0092

Figure 1: Asymptotic Relative Efficiency of the QMLE and RKs

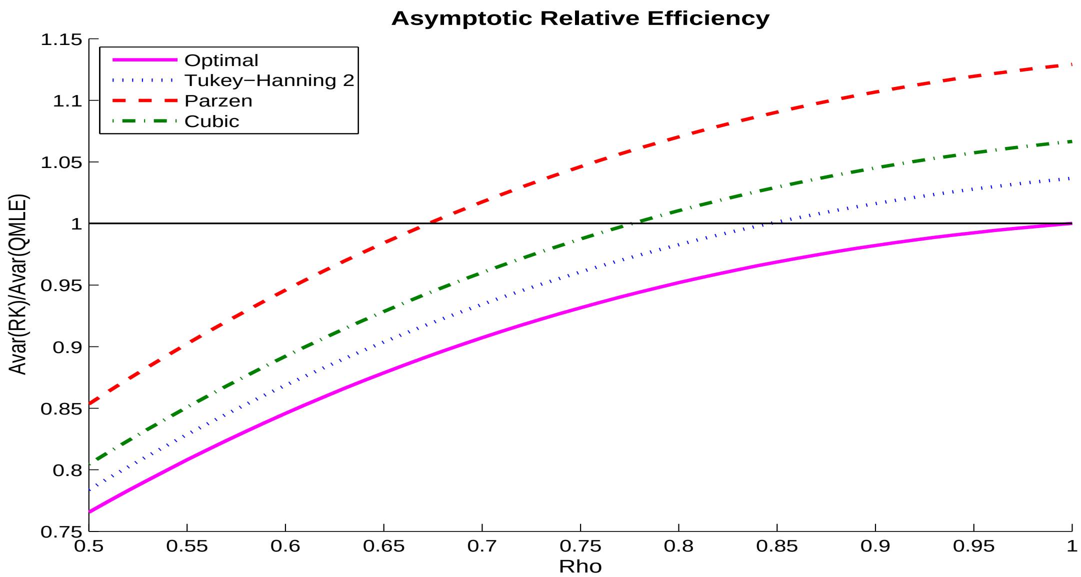
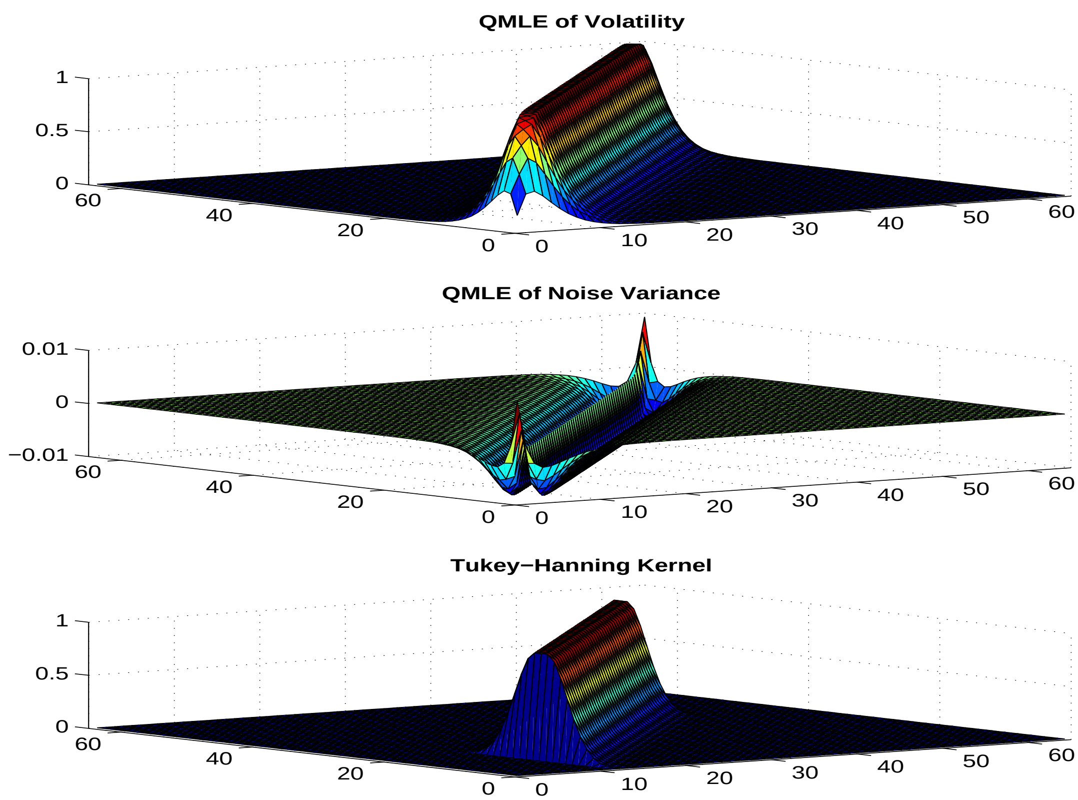
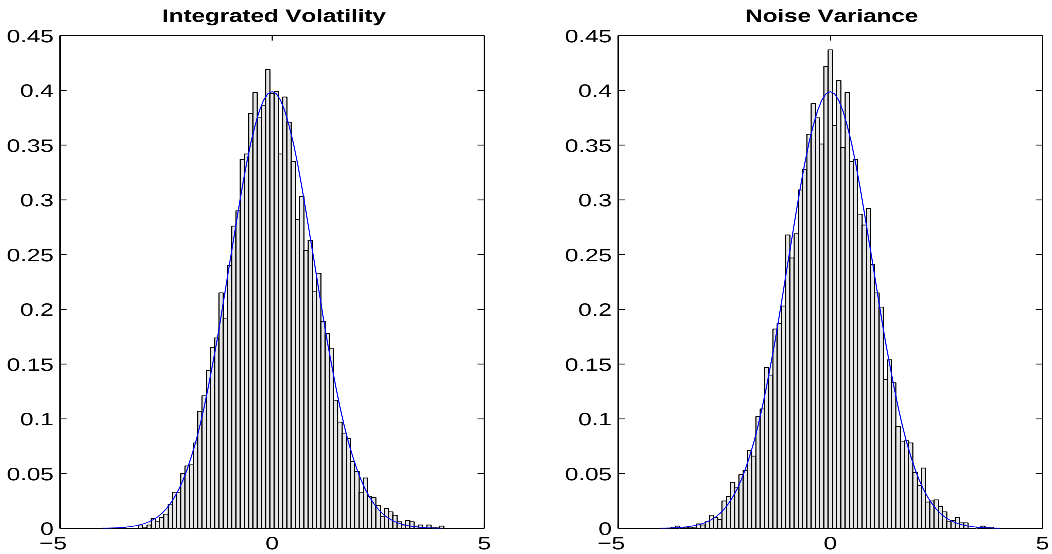


Figure 2: Weighting Matrices of the QMLE and the RK



Note: We plot the two weighting matrices against their column and row indices. The parameters are $\sigma^2 = 0.1$, $a = 0.005$, $n = 65$ and $T = 1/252$. The bandwidth $H = 5.74a\sqrt{n/\sigma^2 T}$.

Figure 3: Histograms of the Standardized Estimates



Note: We plot the histograms of the standardized estimates. The true noise variance is 0.005^2. The simulations include random intervals and volatility jumps. The density of the asymptotic distribution is also plotted.

Figure 4: The EUR/USD Future: A Case Study