6-1-2005

A Two-Stage Realized Volatility Approach to the Estimation for Diffusion Processes from Discrete Observations

Peter C.B. Phillips

Jun Yu

A TWO-STAGE REALIZED VOLATILITY APPROACH TO THE ESTIMATION FOR DIFFUSION PROCESSES FROM DISCRETE OBSERVATIONS

By

Peter C.B. Phillips and Jun Yu

June 2005

COWLES FOUNDATION DISCUSSION PAPER NO. 1523

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

A Two-Stage Realized Volatility Approach to the Estimation for Diffusion Processes from Discrete Observations*

Peter C. B. Phillips[†], Jun Yu[‡]

February 28, 2005

Abstract

This paper motivates and introduces a two-stage method for estimating diffusion processes based on discretely sampled observations. In the first stage we make use of the feasible central limit theory for realized volatility, as recently developed in Barndorff-Nielsen and Shephard (2002), to provide a regression model for estimating the parameters in the diffusion function. In the second stage the in-fill likelihood function is derived by means of the Girsanov theorem and then used to estimate the parameters in the drift function. Consistency and asymptotic distribution theory for these estimates are established in various contexts. The finite sample performance of the proposed method is compared with that of the approximate maximum likelihood method of Aït-Sahalia (2002).

JEL Classification: C13, C22, E43, G13
Keywords: Maximum likelihood, Girsnov theorem, Discrete sampling, Continuous record, Realized volatility.

---

*Phillips gratefully acknowledges support from a Kelly Fellowship at the University of Auckland Business School and from the NSF under Grant No. SES 04-142254. Yu gratefully acknowledge financial support from the Wharton-SMU Research Centre at Singapore Management University.

[†]Peter C. B. Phillips, Cowles Foundation for Research in Economics, Yale University, University of Auckland and University of York; email: peter.phillips@yale.edu

[‡]School of Economics and Social Science, Singapore Management University, 469 Bukit Timah Road, Singapore 259756; email: yujun@smu.edu.sg.

1 Introduction

For many years, continuous time models have enjoyed a great deal of success in finance (Merton, 1990) and more generally in economics (Dixit, 1993). Correspondingly, there has been growing interest in estimating continuous systems using econometric methods.

Many models used in finance for modelling asset prices can be written in terms of a diffusion process as

$$dX_t = \mu(X_t;\theta_1)dt + \sigma(X_t;\theta_2)dB_t, \tag{1}$$

where B_t is a standard Brownian motion, $\sigma(X_t;\theta_2)$ is a known diffusion function, $\mu(X_t;\theta_1)$ is a known drift function, and $\theta = (\theta_1, \theta_2)'$ is a vector of $k_1 + k_2$ unknown parameters. Note that we isolate the vector of parameters θ_2 in the diffusion function from θ_1 for reasons which will be clear below. The attractions of the Ito calculus make it easy to work with processes generated by diffusions like (1) and as a result these processes have been used widely in finance to model asset prices, including stock prices, interest rates, and exchange rates.

From an econometric standpoint, the estimation problem is to estimate θ from observed data which are typically recorded discretely at $(\Delta, 2\Delta, \cdots, n_\Delta\Delta(\equiv T))$ over a certain time interval $[0, T]$, where Δ is the sampling interval and T is the time span of the data. For example, if X_t is recorded as the annualized interest rate and observed monthly (weekly or daily), we have $\Delta = 1/12$ $(1/52$ or $1/250)$. Typically T can be as large as 50 for US Treasury Bills, but is generally much smaller for data from swap markets. Also note that due to time-of-day effects and possibly other market microstructure frictions, it is commonly believed that intra-day data do not follow diffusion models such as (1). As a result, daily and lower frequencies are most frequently used to estimate continuous time models. However, Barndorff-Nielsen and Shephard (2002) and Bollerslev and Zhou (2002) recently showed how to use information from intra-day data to estimate continuous time stochastic volatility models.

A large class of estimation methods is based on the likelihood function derived from the transition probability density of discrete sampling and then resorts to long span asymptotic theory (ie $T \to \infty$). Except for a few cases, the transition probability density does not have a closed form expression and hence the exact maximum likelihood (ML) method based on the likelihood function for the discretely sampled data is not directly available. In the financial econometrics literature, interest in obtaining estimators which approximate or approach ML estimators has been growing, in view of the natural attractiveness of maximum likelihood and its asymptotic properties. Several alternative methods of this type have been developed in recent years.

The main purpose of the present paper is to propose an alternative method of estimating diffusion processes of the form given by model (1) from discrete observations and to establish asymptotic properties by resorting to both the long span (ie $T \to \infty$) and in-fill asymptotics (ie $\Delta \to 0$). The estimation procedure involves two steps. In the first step, we propose to use a quadratic variation type estimator of θ_2. In the second step, an approximate in-fill likelihood function is maximized to obtain a ML estimator of θ_1. This method is not dependent on finding an appropriate auxiliary model, and does not require simulations, nor polynomial expansions. Furthermore, it decomposes the optimization problem into two smaller scale optimization problems. Hence, it is easy to implement and computationally more attractive relative to many other existing methods. The approach also appears to work well in finite samples.

The paper is organized as follows. In Section 2 we review the literature on the ML estimation of diffusion processes and motivate our approach. Section 3 introduces the new method and Section 4 derives the asymptotic properties of the estimates. Section 5 presents some Monte Carlo evidence and Section 6 concludes. Proofs are provided in the Appendix.

2 Literature Review and Motivation

2.1 Literature Review

2.1.1 Transition probability density based approaches

As explained above, a large class of estimation methods is based on the likelihood function derived from the transition probability density of the discretely sampled data. Suppose $p(X_{i\Delta}|X_{(i-1)\Delta}, \theta)$ is the transition probability density. The Markov property of model (1) implies the following log-likelihood function for the discrete sample

$$\ell_{TD}(\theta) = \sum_{i=2}^{n_\Delta} \log(p(X_{i\Delta}|X_{(i-1)\Delta}, \theta)). \tag{2}$$

Under regular conditions, the resulting estimator is consistent, asymptotically normally distributed and asymptotically efficient (Billingsley, 1961). Unfortunately, except for a few cases, the transition density does not have a closed form expression and hence the exact ML method based on the likelihood function of the discrete sample is not a practical procedure. In the financial econometrics literature, interest in finding estimators that approach ML estimators in some quantifiable sense has been growing and many alternative methods have been developed in recent years. For example, Lo (1987) suggested calculating the transition probability density by solving a partial differential equation numerically. Pedersen (1995) and Brandt and Santa-Clara (2002) advocate an approach which calculates the transition probability density using simulation with some auxiliary points between each pair of consecutive observations introduced. This method is also closely related to the Bayesian MCMC method proposed by Elerian, Chib and Shephard (2001) and Eraker (2001). As an important alternative to these numerical and simulated ML methods, Aït-Sahalia (2002) proposed to approximate the transition probability density of diffusions using analytical expansions via Hermite polynomials. Aït-Sahalia (1999) implemented the approximate ML methods and documents its good performance. Apart from these likelihood-based approaches, numerous alternative methods are available. We simply refer readers to the book by Prakasa Rao (1999a) for a review of many alternative approaches.

2.1.2 Approaches based on realized volatility and in-fill likelihood

When the transition probability density does not have a closed form expression but X_t is observed continuously over $[0, T]$, an alternative method can be used to estimate the diffusion models. We now review it in detail.

When the diffusion term is known (ie $\sigma(X_t; \theta_2) = \sigma(X_t)$) and so does not depend on any unknown parameters, one can construct the exact continuous record log-likelihood via the Girsanov theorem (e.g., Liptser and Shiryaev, 2000) as follows.

$$\ell_{IF}(\theta_1) = \int_0^T \frac{\mu(X_t; \theta_1)}{\sigma^2(X_t)} dX_t - \frac{1}{2} \int_0^T \frac{\mu^2(X_t; \theta_1)}{\sigma^2(X_t)} dt.$$

Lánska (1979) established the consistency and asymptotic normality of the continuous record ML estimator of θ_1 when $T \to \infty$ under a certain set of regularity conditions.

The assumptions of a known diffusion function and the availability of a continuous time record are not realistic in financial and other applications. Motivated by the fact that the drift and diffusion functions are of different orders (Bandi and Phillips, 2003, 2004), however, we argue that it seems desirable to estimate the diffusion parameters separately from the drift parameters. For example, when $\sigma(X_t;\theta_2) = \theta_2$, i.e., the diffusion function is an unknown constant, a two-stage approach can be used to estimate the model. First, θ_2 can be estimated by the realized volatility function, i.e.,

$$\hat{\theta}_2 = \sqrt{\frac{[X_\Delta]_T}{T}}, \tag{3}$$

where $[X_\Delta]_T = \sum_{i=2}^{n_\Delta}(X_{i\Delta} - X_{(i-1)\Delta})^2$. This is because model (1) implies that

$$(dX_t)^2 = \theta_2^2 dt, \ \forall t,$$

and hence

$$[X]_T = \int_0^T (dX_t)^2 dt = \int_0^T \theta_2^2 dt = T\theta_2^2,$$

where $[X]_T$ is the quadratic variation of X which can be consistently estimated by $[X_\Delta]_T$ as $\Delta \to 0$. As a result, $\hat{\theta}_2$ should be a very reasonable estimate of θ_2 when Δ is small, which is typically the case for interest rate data. Second, the following logarithmic continuous record likelihood function of model (1)

$$\ell_{IF}(\theta_1) = \int_0^T \frac{\mu(X_t;\theta_1)}{\sigma^2(X_t;\hat{\theta}_2)} dX_t - \frac{1}{2}\int_0^T \frac{\mu^2(X_t;\theta_1)}{\sigma^2(X_t;\hat{\theta}_2)} dt.$$

may be approximated by the in-fill likelihood function

$$\ell_{AIF}(\theta_1) = \sum_{i=2}^{n_\Delta} \frac{\mu(X_{(i-1)\Delta};\theta_1)}{\hat{\theta}_2^2}(X_{i\Delta} - X_{(i-1)\Delta}) - \frac{\Delta}{2}\sum_{i=2}^{n_\Delta} \frac{\mu^2(X_{(i-1)\Delta};\theta_1)}{\hat{\theta}_2^2}, \tag{4}$$

which is in turn maximized with respect to θ_1. This approach is closely related to the method proposed by Florens-Zmirou (1989) where a contrast function instead of the logarithmic in-fill likelihood function was used in the second step.

When the diffusion term is only known up to a scalar factor, that is,

$$dX_t = \mu(X_t;\theta_1)dt + \theta_2 f(X_t)dB_t, \tag{5}$$

the above two-stage method is easily modified. First, θ_2^2 can be estimated by

$$\hat{\theta}_2^2 = \frac{[X_\Delta]_T}{\Delta \sum_{i=2}^{n_\Delta} f^2(X_{(i-1)\Delta})}. \tag{6}$$

Second, the following approximate logarithmic in-fill likelihood function can then be maximized with respect to θ_1 (denoting the resulting estimator by $\hat{\theta}_1$)

$$\ell_{AIF}(\theta_1) = \sum_{i=2}^{n_\Delta} \frac{\mu(X_{(i-1)\Delta};\theta_1)}{\hat{\theta}_2^2 f^2(X_{(i-1)\Delta})}(X_{i\Delta} - X_{(i-1)\Delta}) - \frac{\Delta}{2}\sum_{i=2}^{n_\Delta} \frac{\mu^2(X_{(i-1)\Delta};\theta_1)}{\hat{\theta}_2^2 f^2(X_{(i-1)\Delta})}. \tag{7}$$

This method is applicable to many popular interest rate models, including those proposed by Vasicek (1977), Cox et al (1985) (CIR hereafter), and Ahn and Gao (1998). It is also closely related to the method proposed by Yoshida (1992). In particular, instead of using the estimator in (6), Yoshida (1992) used the following estimator for θ_2^2:

$$\tilde{\theta}_2^2 = \frac{1}{T}\sum_{i=2}^{n_\Delta} \frac{(X_{i\Delta} - X_{(i-1)\Delta})^2}{f^2(X_{(i-1)\Delta})}. \tag{8}$$

Also, Yoshida (1992) suggested using an iterative procedure to construct a better estimate of θ_2^2 (denoted by $\tilde{\theta}_2^2$). Under the conditions of $\Delta \to 0$, $T \to \infty$, and $\Delta^2 T \to 0$, Yoshida (1992) derived the limiting normal distribution for $\sqrt{n_\Delta}(\tilde{\theta}_2^2 - \theta_2^2)$ and $\sqrt{T}(\hat{\theta}_1 - \theta_1)$. Since $\sqrt{n_\Delta}/\sqrt{T} = \sqrt{1/\Delta} \to \infty$, the diffusion parameter enjoys a faster rate of convergence.

The restriction on the diffusion term regarding parameter dependence was somewhat "relaxed" in Hutton and Nelson (1986) who based estimation on the following first order condition of the logarithmic quasi-likelihood function:

$$\int_0^T \frac{\partial\mu(X_t;\theta)/\partial\theta}{\sigma^2(X_t;\theta)} dX_t - \frac{1}{2}\int_0^T \frac{\partial\mu^2(X_t;\theta)/\partial\theta}{\sigma^2(X_t;\theta)} dt = 0.$$

Although their model seems to allow for a more flexible diffusion function, it requires that the drift term share the same set of parameters as the diffusion term. This assumption is too restrictive for practical applications. Moreover, although this one-stage estimation approach is easy to implement, the estimation is mainly based on the drift function and hence leads to inferior finite sample properties, as we will show below in the context of a simple example.

2.2 Motivation

Our two-stage method is in line with the methods proposed by Florens-Zmirou (1989) and Yoshida (1992). That is, in the first stage, we estimate the parameters in the diffusion functions based on the realized volatility, a quantity which consistently estimates the quadratic variation under very mild conditions. In the second step, by assuming the diffusion function is known, we derive and approximate the logarithmic in-fill likelihood function. To motivate the two-step approach, we consider two simple examples.

2.2.1 Example 1

In the first example, we consider estimating the following CIR model

$$dX_t = \kappa(\mu - X_t)dt + \sigma\sqrt{X_t}dB_t, \tag{9}$$

using the exact ML method based on the transition probability density and the two-stage method discussed in Section 2.1.

The natural estimator of σ based on realized volatility is

$$\hat{\sigma} = \sqrt{\frac{[X_\Delta]_T}{\Delta\sum_{i=1}^{n_\Delta} X_{(i-1)\Delta}}}. \tag{10}$$

Moreover, since $\theta_1 = (\kappa, \mu)'$, the logarithmic in-fill likelihood is,

$$\sum_{i=1}^{n} \frac{\kappa(\mu - X_{(i-1)\Delta})(X_{(i-1)\Delta} - X_{(i-1)\Delta})}{\hat{\sigma}^2 X_{(i-1)\Delta}} - \frac{\Delta}{2}\sum_{i=1}^{n} \frac{\kappa^2(\mu - X_{(i-1)\Delta})^2}{\hat{\sigma}^2 X_{(i-1)\Delta}}. \tag{11}$$

CIR (1985) showed that the distribution of $X(t+\Delta)$ conditional on $X(t)$ is non-central chi-squared, $\chi^2[2cX(t), 2q+2, 2\lambda(t)]$, where $c = 2\kappa/(\sigma^2(1-e^{-\kappa\Delta})), \lambda(t) = cr(t)e^{-\kappa\Delta}, q = 2\kappa\mu/\sigma^2 - 1$, and the second and third arguments are the degrees of freedom and non-centrality parameters, respectively. This transition probability density is used to calculate the likelihood function and to obtain the exact ML estimates.

Table 1 reports some results obtained from a Monte Carlo study where we compare two estimation methods. We vary both the sampling frequencies and time spans. Note that the parameters and the sampling frequencies are all set to empirically reasonable values. In all cases, the two-stage method performs comparably with the ML method.

Even in the case where very coarsely sampled data ($\Delta = 1/12$) are available, the two-stage method works quite well. In light of Phillips and Yu (2004), the observed bias in the estimates of κ are the result of the near unit root problem. The observation that the two-stage method is not dominated by ML is quite remarkable, as the data generating process is based on the transition probability density on which ML itself is based. An interesting side result to emerge from this simulation is that the two-stage method is able to reduce the finite sample bias and variance in κ in all cases, even though the reductions are small.

Δ	T	Method	$\kappa = 0.3$		$\mu = 0.09$		$\sigma = 0.06$	
			Mean	SD	Mean	SD $\times 100$	Mean	SD $\times 100$
1/12	20	MLE	.5417	.2832	.0898	1.3848	.0603	.2841
		2-STAGE	.5265	.2663	.0898	1.3785	.0597	.2793
1/12	15	MLE	.6350	.3610	.0903	1.8937	.0604	.3232
		2-STAGE	.6133	.3355	.0904	1.9611	.0596	.3198
1/52	20	MLE	.5075	.2582	.0906	1.3467	.0601	.1332
		2-STAGE	.5045	.2552	.0906	1.3470	.0600	.1347
1/52	10	MLE	.7154	.4390	.0925	2.4234	.0601	.2035
		2-STAGE	.7069	.4306	.0924	2.3301	.0600	.2024
1/250	20	MLE	.5268	.2725	.0898	1.3176	.0600	.0617
		2-STAGE	.5260	.2718	.0898	1.3179	.0600	.0634
1/250	10	MLE	.7533	.4737	.0904	1.9306	.0601	.0874
		2-STAGE	.7519	.4714	.0903	1.9283	.0600	.0891

Table 1: Simulation results under the CIR model, $dX_t = 0.3(0.09 - X_t)dt + 0.06\sqrt{X_t}dB_t$, based on 1000 replications. Mean and SD stand for the average and standard deviation across 1000 replications, respectively.

2.2.2 Example 2

The model in the second example is taken from Hutton and Nelson (1986)

$$dX_t = \alpha dt + \alpha dB_t. \tag{12}$$

Although this model is generally not well suited to interest rate data, the feature that the drift and diffusion functions share the same parameter provides a nice framework to investigate the relative performance of the estimation method based on the diffusion only, against that based on the drift only and that based on the drift and diffusion jointly.

The first method is based on the realized volatility and hence only uses the diffusion term to estimate the model. It is easy to show that

$$\hat{\alpha}_1 = \sqrt{\frac{[X_\Delta]_T}{T}}.$$

The second method is based on the transition probability density given by

$$X_{i\Delta}|X_{(i-1)\Delta} \sim N(X_{(i-1)\Delta} + \alpha\Delta, \alpha^2\Delta).$$

Clearly this method uses information both in the drift and diffusion functions. Denote the resulting estimate by $\hat{\alpha}_2$.

The third method was proposed by Hutton and Nelson (1986). It uses mainly information in the drift function and is based on maximization of the following logarithmic quasi-likelihood function

$$\int_0^T \alpha^{-2} dX_t - \int_0^T \alpha^{-1} dt.$$

As a result, the estimate has the following analytical expression:

$$\hat{\alpha}_3 = \frac{X_T}{T}.$$

Table 2 reports results obtained from a Monte Carlo study where we compare the three estimation methods with different sampling frequencies. In all cases, the two-stage method and ML perform much better than QML; and, most remarkably, the two-stage method performs better than ML. Just as in Example 1, the fact that the simple two-stage method outperforms ML in finite samples is surprising. Moreover, the better performance of the first and second methods clearly reflects the order difference in the drift and diffusion functions.

3 A Two-Stage Method

The estimation procedure discussed in Section 2.1 is not directly applicable to general diffusions such as model 1, as it requires either a constant diffusion function or

Δ	T	True value of $\alpha = 0.1$					
		RV		ML		QML	
		Mean	Variance $\times 100$	Mean	Variance $\times 100$	Mean	Variance $\times 100$
1/12	20	.1013	.0224	.1054	.0266	.0954	.469
1/52	20	.1003	.00488	.1013	.00514	.1005	.5121
1/250	20	.1000	.0011	.1002	.0011	.0992	.518

Table 2: Simulation results under $dX_t = 0.1dt + 0.1dB_t$ based on 1000 replications. Mean and variance are calculated across 1000 replications, respectively.

separability of the scalar parameter from the reminder of the diffusion function. As a result, we have to provide a more general two-step procedure to estimate a diffusion process in the form of model (1). In particular, in the first step we propose to estimate the parameters in the diffusion function by using the feasible central limit theorem for realized volatility derived by Barndorff-Nielsen and Shephard (2002).

Assume that X_t is observed at times

$$t = \Delta, 2\Delta, \cdots, M_\Delta\Delta(= \frac{T}{K}), (M_\Delta + 1)\Delta, \cdots, 2M_\Delta\Delta(= \frac{2T}{K}), \cdots, n_\Delta\Delta(= T),$$

where $n_\Delta = KM_\Delta$ with K being a fixed and positive integer, T is the time span of the data, Δ is the sampling frequency, and $M_\Delta = O(n_\Delta)$. This particular construction allows for the non-overlapping K sub-samples

$$((k-1)M_\Delta + 1)\Delta, \cdots, kM_\Delta\Delta, \quad \text{where } k = 1, \cdots, K,$$

so that each sub-sample has M_Δ observations over the interval $((k-1)\frac{T}{K}, k\frac{T}{K}]$. For example, if ten years of weekly observed interest rates are available and we split the data into ten blocks, then $T = 10$, $\Delta = 1/52$, $M_\Delta = 52$, $K = 10$. The total number of observations is 520 and the number of observations contained in each block is 52.

As $\Delta \to 0$, $n_\Delta = \frac{T}{\Delta} \to \infty$ and $M_\Delta \to \infty$, so that

$$\sum_{i=2}^{M_\Delta} (X_{(k-1)M_\Delta + i\Delta} - X_{(k-1)M_\Delta + (i-1)\Delta})^2 \xrightarrow{p} [X]_{k\frac{T}{K}} - [X]_{(k-1)\frac{T}{K}}, \tag{13}$$

and

$$\frac{\sum_{i=2}^{M_\Delta} (X_{(k-1)M_\Delta + i\Delta} - X_{(k-1)M_\Delta + (i-1)\Delta})^2 - ([X]_{k\frac{T}{K}} - [X]_{(k-1)\frac{T}{K}})}{r_k} \xrightarrow{d} N(0,1), \tag{14}$$

$$\frac{\log(\sum_{i=2}^{M_\Delta}(X_{(k-1)M_\Delta+i\Delta}-X_{(k-1)M_\Delta+(i-1)\Delta})^2-\log([X]_{k\frac{T}{K}}-[X]_{(k-1)\frac{T}{K}})}{s_k}\xrightarrow{d}N(0,1), \tag{15}$$

where

$$r_k=\sqrt{\frac{2}{3}\sum_{i=2}^{M_\Delta}(X_{(k-1)M_\Delta+i\Delta}-X_{(k-1)M_\Delta+(i-1)\Delta})^4}$$

and

$$s_k=\min\{\sqrt{\frac{r_k^2}{(\sum_{i=2}^{M_\Delta}(X_{(k-1)M_\Delta+i\Delta}-X_{(k-1)M_\Delta+(i-1)\Delta})^2)^2}},\frac{2}{M_\Delta}\} \tag{16}$$

for $k=1,\cdots,K$. The limit (13) follows by virtue of the definition of quadratic variation, while the central limit theorem (CLT) results (14) and (15) are due to Barndorff-Nielsen and Shephard (2002), where (16) involves a finite sample correction on the asymptotic theory of Barndorff-Nielsen and Shephard (2005).

Based on the CLT (14), θ_2 can be estimated in the first stage by running a (nonlinear) least squares regression of the standardized realized volatility

$$\frac{\sum_{i=2}^{M_\Delta}(X_{(k-1)M_\Delta+i\Delta}-X_{(k-1)M_\Delta+(i-1)\Delta})^2}{r_k} \tag{17}$$

on the standarized diffusion function

$$\begin{aligned}\frac{([X]_{k\frac{T}{K}}-[X]_{(k-1)\frac{T}{K}})}{r_k} &= \frac{\left(\int_{(k-1)\frac{T}{K}}^{k\frac{T}{K}}\sigma^2\left(X_t;\theta_2\right)dt\right)}{r_k} &(18)\\ &\simeq \frac{\sum_{i=2}^{M_\Delta}\sigma^2\left(X_{(k-1)M_\Delta+(i-1)\Delta};\theta_2\right)\Delta}{r_k} &(19)\end{aligned}$$

for $k=1,\cdots,K$. Denote the resulting estimator of θ_2 by $\hat{\theta}_2$. In fact, we can write $\hat{\theta}_2$ as the extremum estimator

$$\hat{\theta}_2=\arg\min_{\theta_2}Q_\Delta\left(\theta_2\right), \tag{20}$$

where

$$Q_\Delta\left(\theta_2\right)=\Delta\sum_{k=1}^{K}\left[\frac{\sum_{i=2}^{M_\Delta}\left\{(X_{(k-1)M_\Delta+i\Delta}-X_{(k-1)M_\Delta+(i-1)\Delta})^2-\sigma^2\left(X_{(k-1)M_\Delta+(i-1)\Delta};\theta_2\right)\Delta\right\}}{r_k}\right]^2.$$

A similar regression in standardized log levels of realized volatility can be run using result (15).

This approach provides a more general estimation procedure than those designed to estimate models with a constant diffusion or a scalar parameter in the diffusion function. Indeed, when $K = 1$, the least squares regression above is equivalent to minimizing the squared difference between the terms given by Equations (17) and (18), which yields exactly the expression of the estimator (6) when the diffusion term is known up to the scalar factor.

In the second stage, the approximate log-likelihood function is maximized with respect to θ_1 (denoting the resulting estimator by $\hat{\theta}_1$)

$$\ell_{AIF}(\theta_1) = \sum_{i=2}^{n_\Delta} \frac{\mu(X_{(i-1)\Delta};\theta_1)}{\sigma^2(X_{(i-1)\Delta};\hat{\theta}_2)}(X_{i\Delta} - X_{(i-1)\Delta}) - \frac{\Delta}{2}\sum_{i=2}^{n_\Delta} \frac{\mu^2(X_{(i-1)\Delta};\theta_1)}{\sigma^2(X_{(i-1)\Delta};\hat{\theta}_2)}. \tag{21}$$

4 Asymptotic Results

The asymptotic theory of a slightly different two-stage estimator in the multivariate case has been obtained in Yoshida (1992) for models whose diffusion term is known up to a constant (matrix) factor, where both infill and long span asymptotics are employed both for the diffusion and drift parameter estimators. In this section we first derive the asymptotic theory for the same class of (scalar) models but only resort to long span asymptotics for the drift parameter asymptotic theory. We then investigate the asymptotic properties of the estimators proposed in Section 3 for model (1) whose diffusion function has a general form.

4.1 Scalar Parameter in the Diffusion Function

Assume the data are generated from the following stochastic differential equation:

$$dX_t = \mu(X_t;\theta_1^*)dt + \theta_2^* f(X_t)dB_t. \tag{22}$$

Denote θ_2^2 by τ and θ_2^{*2} by τ^*. Both $\mu(\cdot;\theta_1)$ and $f(\cdot)$ are time-homogeneous, $\mathcal{B}$-measurable functions on $\mathcal{D} = (l,u)$ with $-\infty \leq l < u \leq \infty$, where $\mathcal{B}$ is the σ-field generated by Borel sets on $\mathcal{D}$. τ is estimated by $\hat{\tau}$ defined by Equation (6); θ_1 is estimated by $\hat{\theta}_1$, the maximizer of Equation (7).

To prove consistency of $\hat{\tau}$ in a diffusion process with a constant diffusion term (ie $f(X_t) = 1$), Florens-Zmirou (1989) assumed $\Delta \to 0$, $T \to \infty$, and $\Delta^2 T \to 0$. The same set of assumptions were employed by Yoshida (1992) to deal with the diffusion process for more general, but still known, $f(X_t)$. In this paper, using the theory of Barndorff-Nielsen and Shephard (2002) we show that the condition of an infinite time span of data (ie $T \to \infty$) is not needed to develop the asymptotic theory for $\hat{\tau}$.

We list the following conditions.

Assumption 1: *Equation* $[X]_t - \tau \int_0^t f^2(X_s)ds = 0$ *has a unique solution at* $\tau^* > 0 \; \forall t > 0$.

Assumption 2: $\inf_{x \in J} f^2(x) > 0$, *where* J *is a compact subset of the range of the process.*

Assumption 3: $\int_0^t \mu^2(X_s; \theta_1)ds > 0 \; \forall t < \infty$.

Remark 4.1: Assumption 1 is an identification condition. Assumption 3 ensures weak convergence of the error process from the Euler approximation to the diffusion process (Jacod and Protter, 1998).

THEOREM 4.1 (Asymptotics of the Diffusion Parameter Estimate): *Suppose Assumptions 1-2 hold,* $\hat{\tau} \xrightarrow{p} \tau^*$ *as* $\Delta \to 0$. *If, in addition, Assumption 3 holds,*

$$\Delta^{-1/2}(\hat{\tau} - \tau^*) \xrightarrow{d} \frac{\sqrt{2}\int_0^T \tau^* f^2(X_s)dW_s}{\int_0^T f^2(X_s)ds}.$$

where W_t *is a Brownian motion which is independent of* X_t.

Remark 4.2: With a different estimate for τ, we improve the results of Yoshida (1992), who derived asymptotic properties of diffusion estimate by assuming $\Delta \to 0$ and $T \to \infty$, by only requiring in Theorem 4.1 that $\Delta \to 0$.

To establish the asymptotic properties of the drift parameter estimate, we follow Yoshida (1992) closely. In particular, we first list the following conditions.

Assumption 4: $\theta_1 \in \mathbf{\Theta}_1$ *where the parameter space* $\mathbf{\Theta}_1 \subset R^{K_1}$ *is a compact set with* $\theta_1^* \in Int(\mathbf{\Theta}_1)$.

Assumption 5: *Both* $\mu(\cdot; \theta_1)$ *and* $f(\cdot)$ *functions are twice continuously differentiable. As a result, for any compact subset* J *of the range of the process, we have the following two conditions:*

(i) (Lipschitz condition) There exists a constant L_1 *so that*

$$|\mu(x; \theta_1^*) - \mu(y; \theta_1^*)| + \theta_2^*|f(x) - f(y)| \leq L_1|x - y|,$$

for all x and y in J.

(ii) (Growth condition) There exists a constant L_2 so that

$$|\mu(x;\theta_1^*)| + \theta_2^*|f(x)| \le L_2|1+x|,$$

for all x and y in J.

Assumption 6: *Define the scale measure of X_t by*

$$s(x;\theta) = \exp\bigg(-2\int_c^x \frac{\mu(y;\theta_1)}{\tau f^2(y)} dy \bigg),$$

where c is a generic constant. We assume the following conditions hold

$$\int_c^u s(x;\theta)dx = \int_l^c s(x;\theta)dx = \infty,$$

and

$$\int_l^u \frac{1}{s(x;\theta)\tau f^2(x)} dx = A(\theta) < \infty.$$

Assumption 7: *For arbitrary $p \ge 0$,*

$$\sup_t E(|X_t|^p) < \infty.$$

Assumption 8: *Define the following function*

$$\theta_1 \to Y(\theta_1;\tau^*) = \int \frac{\mu(x,\theta_1)}{\tau^* f^2(x)}(\mu(x,\theta_1^*) - \frac{1}{2}\mu(x,\theta_1))\pi_\theta(dx)$$

and assume function $Y(\cdot;\tau^*)$ has the unique maximum at $\theta_1 = \theta_1^*$, where π_θ is defined in Remark 4.4.

Assumption 9: *For fixed θ_1, the derivatives $\partial^l \mu(x;\theta_1)/\partial x^l$ and $\partial^l f(x)/\partial x^l$ ($l = 1,2$) exist and they are continuous in x. For fixed x, $\partial^l \mu(x;\theta_1)/\partial \theta_1^l$ exist. Moreover,*

$$|\partial^l \mu(x;\theta_1)/\partial x^l|, |\partial^l f(x)/\partial x^l|, |\partial^l \mu(x;\theta_1)/\partial \theta_1^l| \le C(1+|x|)^C,$$

for $l = 0,1,2$.

Assumption 10: *The matrix*

$$\Phi = \int \frac{\partial \mu(x;\theta_1^*)}{\partial \theta_1^\top}(\tau^* f^2(x))^{-1}\frac{\partial \mu(x;\theta_1^*)}{\partial \theta_1}\pi_\theta(dx) \tag{23}$$

is positive definite.

Remark 4.3: Under Assumption 5, these exists a solution process for the stochastic differential equation and the solution is unique.

Remark 4.4: Under Assumption 6, the process X_t is ergodic with an invariant probability measure that has density

$$\pi_\theta(x) = \frac{1}{A(\theta)s(x;\theta)\tau f^2(x)},$$

for $x \in (l, u)$ with respect to Lebesque measure on (l, u), where $A(\theta)$ and $s(x; \theta)$ are defined in Assumption 6. We further assume that $X_0 \sim \pi_{\theta^*}$ so that X_t is a stationary process with $X_t \sim \pi_{\theta^*}$.

THEOREM 4.2 (Asymptotics of the Drift Parameter Estimates): *Let* $\hat{\theta}_1 = \text{argmax}_{\,\theta_1 \in \mathbf{\Theta}_1} T^{-1} \log \ell_{AIF}(\theta_1)$ *with* $\ell_{AIF}(\theta_1)$ *given by Equation (7). Suppose Assumptions 1-10 hold,* $\hat{\theta}_1 \xrightarrow{p} \theta_1^*$ *as* $\Delta \to 0$ *and* $T \to \infty$*. If, in addition,* $\Delta^2 T \to 0$*,*

$$T^{1/2}(\hat{\theta}_1 - \theta_1^*) \xrightarrow{d} N(0, \Phi^{-1}),$$

where Φ *is given in Equation (23).*

4.2 General Diffusions

Suppose data are generated from the following stochastic differential equation

$$dX_t = \mu(X_t; \theta_1^*)dt + \sigma(X_t; \theta_2^*)dB_t, \tag{24}$$

where $\theta_1 \in \mathbf{\Theta}_1 \subset R^{K_1}$ and $\theta_2 \in \mathbf{\Theta}_2 \subset R^{K_2}$. Both $\mu(\cdot, \theta_1)$ and $\sigma(\cdot; \theta_2)$ are time-homogeneous, $\mathcal{B}$-measurable functions on $\mathcal{D} = (l, u)$ with $-\infty \leq l < u \leq \infty$, where $\mathcal{B}$ is the σ-field generated by Borel sets on $\mathcal{D}$. θ_2 is estimated by regressing (17) on (18), giving the extremum estimator (20); θ_1 is estimated by $\hat{\theta}_1$, defined by Equation (21).

As in the scalar factor parameter case, we show that an infinite time span (ie $T \to \infty$) is not needed to develop the asymptotic theory for $\hat{\theta}_2$.

Some additional assumptions are required, given the nonlinear dependence of the diffusion $\sigma(X_t; \theta_2)$ on θ_2**.** Also we have to modify some earlier Assumptions listed in Section 4.1.

Assumption $1'$**:** *The equation*

$$[X]_t - \int_0^t \sigma^2(X_s; \theta_2)ds = \int_0^t \sigma^2(X_s; \theta_2^*)ds - \int_0^t \sigma^2(X_s; \theta_2)ds = 0 \tag{25}$$

has a unique solution at θ_2^*, $\forall t > 0$.

Assumption $2'$**:** $\inf_{x\in J}\sigma^2(x;\theta_2^*) > 0$, *where* J *is a compact subset of the range of the process.*

Assumption $4'$: $\theta_1 \in \boldsymbol{\Theta}_1$, $\theta_2 \in \Theta_2$, *where parameter spaces* $\boldsymbol{\Theta}_1 \subset R^{k_1}$ *and* $\Theta_2 \subset R^{k_2}$ *are compact set with* $\theta_1^* \in Int(\boldsymbol{\Theta}_1)$ *and* $\theta_2^* \in Int(\Theta_2)$.

Assumption $5'$: *Both* $\mu(\cdot;\theta_1)$ *and* $\sigma(\cdot;\theta_2)$ *functions are twice continuously differentiable. As a result, for any compact subset* J *of the range of the process, we have the following two conditions:*

(i) (Lipschitz condition) There exists a constant L_1 *so that*

$$|\mu(x;\theta_1^*) - \mu(y;\theta_1^*)| + |\sigma(x;\theta_2^*) - \sigma(y;\theta_2^*)| \le L_1|x-y|,$$

for all x *and* y *in* J.

(ii) (Growth condition) There exists a constant L_2 *so that*

$$|\mu(x;\theta_1^*)| + |\sigma(x;\theta_2^*)| \le L_2|1+x|,$$

for all x *and* y *in* J.

Assumption $6'$**:** *Define the scale measure of* X_t *by*

$$s(x;\theta) = \exp\left(-2\int_c^x \frac{\mu(y;\theta_1)}{\sigma^2(y;\theta_2)}dy\right),$$

where c is a generic constant. We assume the following condition holds

$$\int_c^u s(x;\theta)dx = \int_l^c s(x;\theta)dx = \infty,$$

and

$$\int_l^u \frac{1}{s(x;\theta)\sigma^2(x;\theta_2)}dx = A(\theta) < \infty.$$

Assumption $8'$: *Define the following function*

$$\theta_1 \to Y(\theta_1;\theta_2^*) = \int \frac{\mu(x,\theta_1)}{\sigma^2(x;\theta_2^*)}(\mu(x,\theta_1^*) - \frac{1}{2}\mu(x,\theta_1))\pi_{\theta^*}(dx)$$

and assume $Y(\cdot;\theta_2^*)$ *has the unique maximum at* $\theta_1 = \theta_1^*$.

Assumption $9'$**:** *For fixed* θ_1*, the derivatives* $\partial^l \mu(x;\theta_1)/\partial x^l$ *and* $\partial^l \sigma(x;\theta_2)/\partial x^l$ *(*$l = 1,2$*) exist and they are continuous in* x*. For fixed* x*,* $\partial^l \mu(x;\theta_1)/\partial \theta_1^l$ *and* $\partial^l \sigma(x;\theta_2)/\partial \theta_2^l$ *exist. Moreover,*

$$|\partial^l \mu(x;\theta_1)/\partial x^l|, |\partial^l \sigma(x;\theta_2)/\partial x^l|, |\partial^l \mu(x;\theta_1)/\partial \theta_1^l|, |\partial^l \sigma(x;\theta_2)/\partial \theta_2^l| \leq C(1+|x|)^C,$$

for $l = 0,1,2$.

Assumption $10'$**:** *The matrices*

$$\Phi_1 = \int \frac{\partial \mu(x;\theta_1^*)}{\partial \theta_1} \sigma^{-2}(x;\theta_2^*) \frac{\partial \mu(x;\theta_1^*)}{\partial \theta_1'} \pi_\theta(dx) \tag{26}$$

$$\text{and} \int_0^t \frac{\partial \sigma^2(X_s;\theta_2^*)}{\partial \theta_2} \frac{\partial \sigma^2(X_s;\theta_2^*)}{\partial \theta_2'} ds$$

are positive definite and $\int_0^t \sigma^4(X_s;\theta_2^*)\, ds > 0$ *for all* $t > 0$.

THEOREM 4.3 (Asymptotics of the Diffusion Parameter Estimate): *Suppose Assumptions* $1'$*-*$10'$ *hold. Then,* $\hat{\theta}_2 \xrightarrow{p} \theta_2^*$ *as* $\Delta \to 0$ *and*

$$\Delta^{-1/2}\left(\hat{\theta}_2 - \theta_2^*\right) \xrightarrow{d} \left[\sum_{k=1}^{K} \frac{\int_{(k-1)\frac{T}{K}}^{k\frac{T}{K}} \frac{\partial \sigma^2(X_s;\theta_2^*)}{\partial \theta_2} \frac{\partial \sigma^2(X_s;\theta_2^*)}{\partial \theta_2'} ds}{\int_{(k-1)\frac{T}{K}}^{k\frac{T}{K}} \sigma^4(X_s;\theta_2^*)\, ds}\right]^{-1}$$
$$\times \left[\sum_{k=1}^{K} \frac{\sqrt{2}\int_{(k-1)\frac{T}{K}}^{k\frac{T}{K}} \frac{\partial \sigma^2(X_s;\theta_2^*)}{\partial \theta_2} \sigma^2(X_s;\theta_2^*)\, dW_s}{2\int_{(k-1)\frac{T}{K}}^{k\frac{T}{K}} \sigma^4(X_s;\theta_2^*)\, ds}\right],$$

where W_t *is a Brownian motion which is independent of* X_t*.*

THEOREM 4.4 (Asymptotics of the Drift Parameter Estimate): Let $\hat{\theta}_1 = \operatorname{argmax} T^{-1} \log \ell_{AIF}(\theta_1)$ *with* $\ell_{AIF}(\theta_1)$ *given by Equation (21). Suppose Assumptions* $1' - 10'$ *hold, then* $\hat{\theta}_1 \xrightarrow{p} \theta_1^*$ *as* $\Delta \to 0$ *and* $T \to \infty$*. If, in addition,* $\Delta^2 T \to 0$,

$$T^{1/2}(\hat{\theta}_1 - \theta_1^*) \xrightarrow{d} N(0, \Phi_1^{-1}),$$

where Φ_1 *is given in Equation (26).*

5 Monte Carlo Results

To examine the performance of the proposed procedure, we estimate the following model for short-term interest rates due to Chan et al. (CKLS hereafter) (1992),

$$dX_t = \kappa(\mu - X_t)dt + \sigma X_t^{\gamma} dB_t, \tag{27}$$

with $\kappa = 0.6, \mu = 0.09, \sigma = 0.06, \gamma = 0.5$. We choose $\gamma = 0.5$ so that the true model becomes a CIR model which enables an exact data simulation. The parameters are estimated from 10 years of daily data (2500 observations). The experiment is replicated 1000 times to get the means and standard errors for each estimate. Two estimation methods are employed to estimate the model: the approximate ML method of Aït-Sahalia (2002) and the proposed two-stage method.[1] The results are reported in Table 3.

To use the two-stage method, the number of subsamples has to be chosen. Since there are two parameters in the diffusion term, $K = 1$ is not adequate. Although $K = 2$ may seem a natural choice, the simulation results suggest that larger values of K are better and the performance of the procedure improves substantially when $K \geq 10$. The estimation of parameters in the drift function does not seem to be dependent on the diffusion parameters in any critical way. This suggests that in all cases the quadratic variations are well estimated.

6 Conclusion

This paper proposes a two-stage method to estimate diffusion processes in a general form. In the first stage the realized volatility calculated from a sequence of split samples is regressed on the corresponding quadratic variation in order to estimate all the parameters in the diffusion function. Then, conditional on the resulting consistent estimate of the diffusion, the in-fill likelihood function approximation of the diffusion process can be readily constructed. The resulting discrete approximation produces estimates of all the parameters in the drift function. Monte Carlo simulations show

[1] Although the asymptotic theory has been developed for the standardized realized volatility in the present paper, the finite sample performance often improves for the regression based on the log realized volatility. As a result, the Monte Carlo results reported in Table 3 are based on Equation (15).

		AML	Two-Stage Method			
			K=2	K=10	K=20	K=50
γ	Mean	0.4901	0.5326	0.4992	0.4954	0.4927
(=0.5)	SD	0.1044	0.5117	0.1295	0.1136	0.1104
σ	Mean	0.0604	0.1562	0.0628	0.0612	0.0603
(=0.06)	SD	0.0157	0.4754	0.0235	0.0177	0.0167
κ	Mean	1.0716	1.0697	1.0729	1.0730	1.0726
(=0.6)	SD	0.5447	0.5399	0.5417	0.5417	0.5415
μ	Mean	0.0901	0.0902	0.0902	0.0902	0.0902
(=0.09)	SD	0.0097	0.0094	0.0094	0.0094	0.0094

Table 3: Simulation results under the CKLS model, $dX_t = \kappa(\mu - X_t)dt + \sigma X_t^\gamma dB_t$, based on 1000 samples of 2500 daily observations. Mean and SD stand for the average and standard deviation across 1000 replications, respectively.

that the finite sample performance of the proposed method is very satisfactory and as good as conventional maximum likelihood even when the discrete likelihood can be obtained. One advantage of the proposed method is that a larger scale optimization problem is decomposed into two smaller scale optimization problems. Although, like other extreme estimators, our method tends to over estimate the mean reversion parameter, κ, the numerical attractability of our method makes it an ideal initial estimate for the jackknife method of Phillips and Yu (2004) to reduce the finite sample bias in κ.

The approach can be readily extended to the multi-dimensional case. Both implementation and asymptotic theory only need trivial modifications. Since the method separates estimation of the drift and diffusion functions, it may be a desirable method to use when the drift but not the diffusion involves certain market microstructure features.

7 Appendix

Proof of Theorem 4.1: It is known that all diffusion-type processes are semi-martingales (Prakasa Rao, 1999b). As a result, when $\Delta \to 0$,

$$[X_\Delta]_T \xrightarrow{p} [X]_T = \tau^* \int_0^T f^2(X_s)ds,$$

where the convergence follows from the theory of quadratic variation for semi-martingales and the equality follows from Assumption 1.

By Assumption 3, we have

$$\hat{\tau} = \frac{[X_\Delta]_T}{\sum_{i=1}^{n_\Delta} f^2(X_{(i-1)\Delta})} \xrightarrow{p} \frac{[X]_T}{\int_0^T f^2(X_s)ds} = \tau^*.$$

This proves the first part of Theorem 1.

Since X_t is a semi-martingale, by Ito's lemma for semi-martinagles (Prakasa Rao, 1999b) we have

$$X_T^2 = [X]_T + 2\int_0^T X_{s-}dX_{s-}.$$

Following Theorem 1 of Barndorff-Nielsen and Shephard (2002) we have

$$\Delta^{-1/2}([X_\Delta]_T - [X]_T) \xrightarrow{d} \tau^*\sqrt{2}\int_0^T f^2(X_s)dW_s, \tag{28}$$

where W_t is a Brownian motion which is independent of X_t. Hence,

$$\begin{aligned} &\Delta^{-1/2}(\hat{\tau} - \tau^*) \\ &= \Delta^{-1/2}\left(\frac{[X_\Delta]_T}{\sum_{i=1}^{n_\Delta} f^2(X_{(i-1)\Delta})} - \frac{[X]_T}{\int_0^T f^2(X_s)ds}\right) \\ &= \Delta^{-1/2}\frac{1}{\int_0^T f^2(X_s)ds}\left(\frac{\int_0^T f^2(X_s)ds}{\sum_{i=1}^{n_\Delta} f^2(X_{(i-1)\Delta})}[X_\Delta]_T - [X]_T\right). \end{aligned} \tag{29}$$

By Assumption 3,

$$\frac{\sum_{i=1}^{n_\Delta} f^2(X_{(i-1)\Delta})}{\int_0^T f^2(X_s)ds} \xrightarrow{p} 1.$$

By Slutsky's theorem, Equations (28) and (29) imply that

$$\Delta^{-1/2}(\hat{\tau} - \tau^*) \xrightarrow{d} \tau^* \frac{\sqrt{2}\int_0^T f^2(X_s)dW_s}{\int_0^T f^2(X_s)ds}. \tag{30}$$

This completes the proof of Theorem 1. ∎

Proof of Theorem 4.2:

Obviously, the proposed drift estimator is in the class of extremum estimators. Hence, one can prove consistency by checking sufficient conditions for extremum estimation problems. It is convenient here to check the conditions given in Newey and McFadden (1994, p.2121), namely, compactness, continuity, uniform convergence, and identifiability.

Compactness of $\boldsymbol{\Theta}$, continuity of $T^{-1}\log \ell_{AIF}(\theta_1;\hat{\tau})$ and the identification condition are assured by Assumption 1, Assumption 9 and Assumption 8, respectively. The uniform convergence of $T^{-1}\log \ell_{AIF}(\theta_1)$ to $Y(\theta_1, \tau^*)$ follows from Proposition 1, Lemma 1 and Lemma 2 in Yoshida (1992). Hence the first part of the theorem is proved.

To show asymptotic normality, we follow Yoshida (1992) by obtaining the weak convergence of the likelihood ratio random field,

$$Z_{\Delta,n_\Delta}(\tau,u) = \ell_{AIF}(\theta_1^* + T^{-1/2}u;\hat{\tau})/\ell_{AIF}(\theta_1^*;\hat{\tau}).$$

Under the listed conditions, Yoshida (1992) showed that

$$\begin{aligned} \log Z_{\Delta,n_\Delta}(\tau^*,u) \;=\;& u^\top T^{-1/2}\sum_{i=1}^{n_\Delta}\frac{\partial\mu(x;\theta_1^*)}{\partial\theta_1}\frac{1}{\tau^* f^2(X_{(i-1)\Delta})}\int_{(i-1)\Delta}^{i\Delta}\tau^* f(x)dW_t \\ & -\frac{1}{2}u^\top\Phi u + \rho_{\Delta,n}(u), \end{aligned} \tag{31}$$

where $\rho_{\Delta,n}(u) \xrightarrow{p} 0$ and Φ is defined in Equation (23).

From Theorem 1, we have, $\forall\eta > 0$, that there exists a Δ and a positive number c_1 such that

$$P(\Delta^{-1/2}(\hat{\tau} - \tau^*) > c_1) < \eta/2.$$

Let $\hat{\tau} = \tau^* + \Delta^{1/2} M$ and we have, $\forall \epsilon > 0$

$$
\begin{aligned}
& P(|\log Z_{\Delta,n_\Delta}(\hat{\tau}, u) - \log Z_{\Delta,n_\Delta}(\tau^*, u)| > \epsilon) \\
= \;& P(\Delta^{-1/2}(\hat{\tau} - \tau^*) > c_1) + P(\sup_{|M|\le c_1} |\log Z_{\Delta,n}(\hat{\tau}, u) - \log Z_{\Delta,n_\Delta}(\tau^*, u)| > \epsilon) \\
< \;& \eta. \qquad (32)
\end{aligned}
$$

Combining equations (31) and (32) proves Proposition 4 of Yoshida (1992). Similarly, we can obtain Propositions 5 and 6 of Yoshida (1992) based on $\hat{\tau}$. The week convergence of the likelihood random field follows these propositions. In particular,

$$T^{1/2}(\hat{\theta}_1 - \theta_1^*) \xrightarrow{d} N(0, \Phi).$$

This completes the proof of Theorem 4.2. ■

Proof of Theorem 4.3:

The argument is briefly sketched here. We consider the case where the estimate $\hat{\theta}_2$ is obtained from the extremum estimation problem (20), viz.,

$$\hat{\theta}_2 = \arg\min_{\theta_2} Q_\Delta(\theta_2),$$

where

$$Q_\Delta(\theta_2) = \Delta \sum_{k=1}^{K} \left[\frac{\sum_{i=2}^{M_\Delta} \left\{ (X_{(k-1)M_\Delta + i\Delta} - X_{(k-1)M_\Delta+(i-1)\Delta})^2 - \sigma^2\left(X_{(k-1)M_\Delta+(i-1)\Delta}; \theta_2\right)\Delta \right\}}{r_k} \right]^2.$$

A similar argument can be employed in the case where standardized log levels of realized volatility are used in the regression based on the CLT result (15).

Observe that, as $\Delta \to 0$,

$$Q_\Delta(\theta_2) \xrightarrow{p} Q(\theta_2) = \sum_{k=1}^{K} \frac{\left\{ \int_{(k-1)\frac{T}{K}}^{k\frac{T}{K}} \sigma^2(X_s; \theta_2^*)\, ds - \int_{(k-1)\frac{T}{K}}^{k\frac{T}{K}} \sigma^2(X_s; \theta_2)\, ds \right\}^2}{2\int_{(k-1)\frac{T}{K}}^{k\frac{T}{K}} \sigma^4(X_s; \theta_2^*)\, ds},$$

uniformly in θ_2, since

$$\sum_{i=2}^{M_\Delta} (X_{(k-1)M_\Delta+i\Delta} - X_{(k-1)M_\Delta+(i-1)\Delta})^2 \xrightarrow{p} [X]_{k\frac{T}{K}} - [X]_{(k-1)\frac{T}{K}} = \int_{(k-1)\frac{T}{K}}^{k\frac{T}{K}} \sigma^2(X_s; \theta_2^*)\, ds,$$

and

$$\sum_{i=2}^{M_\Delta} \sigma^2 \left(X_{(k-1)M_\Delta+(i-1)\Delta};\theta_2\right)\Delta \xrightarrow{p} \int_{(k-1)\frac{T}{K}}^{k\frac{T}{K}} \sigma^2 \left(X_s;\theta_2\right) ds, \tag{33}$$

uniformly in $\theta_2 \in \Theta_2$ in view of the compactness of Θ_2 and the smoothness of $\sigma^2 \left(X_s;\theta_2\right)$. Next,

$$\frac{r_k^2}{\Delta} = \frac{2}{3\Delta}\sum_{i=2}^{M_\Delta}(X_{(k-1)M_\Delta+i\Delta} - X_{(k-1)M_\Delta+(i-1)\Delta})^4 \xrightarrow{p} 2\int_{(k-1)\frac{T}{K}}^{k\frac{T}{K}} \sigma^4 \left(X_s;\theta_2^*\right) ds, \tag{34}$$

as in Barndorff-Nielsen and Shephard (2002, hereafter BNS). Thus, since $Q\left(\theta_2\right)$ is minimized for $\theta_2 = \theta_2^*$ in view of (25), we have $\hat{\theta}_2 \xrightarrow{p} \theta_2^*$ by a standard extremum estimator argument.

Next, by a Taylor series argument under the stated smoothness and positive definiteness assumptions, we have

$$\begin{aligned}\Delta^{-1/2}\left(\hat{\theta}_2 - \theta_2^*\right) &= \left[\frac{1}{\Delta}\frac{\partial^2 Q_\Delta\left(\tilde{\theta}_2\right)}{\partial\theta_2\partial\theta_2'}\right]^{-1}\left[\frac{1}{\Delta^{3/2}}\frac{\partial Q_\Delta\left(\theta_2^*\right)}{\partial\theta_2}\right] \\ &\sim \left[\frac{1}{\Delta}\frac{\partial^2 Q_\Delta\left(\theta_2^*\right)}{\partial\theta_2\partial\theta_2'}\right]^{-1}\left[\frac{1}{\Delta^{3/2}}\frac{\partial Q_\Delta\left(\theta_2^*\right)}{\partial\theta_2}\right], \end{aligned} \tag{35}$$

where $\tilde{\theta}_2$ is on the line segment connecting $\hat{\theta}_2$ to θ_2^* and thus satisfies $\tilde{\theta}_2 \to_p \theta_2^*$. Setting

$$g_i^* = g\left(X_{(k-1)M_\Delta+(i-1)\Delta};\theta_2^*\right) = \frac{\partial\sigma^2\left(X_{(k-1)M_\Delta+(i-1)\Delta};\theta_2^*\right)}{\partial\theta_2},$$

we get

$$\begin{aligned}&\frac{1}{\Delta^{3/2}}\frac{\partial Q_\Delta\left(\theta_2^*\right)}{\partial\theta_2} \\ =\; & -\frac{2}{\Delta^{3/2}}\sum_{k=1}^{K}\frac{\sum_{i=2}^{M_\Delta} g_i^*\Delta\left\{(X_{(k-1)M_\Delta+i\Delta} - X_{(k-1)M_\Delta+(i-1)\Delta})^2 - \sigma^2\left(X_{(k-1)M_\Delta+(i-1)\Delta};\theta_2\right)\Delta\right\}}{\frac{r_k^2}{\Delta}} \\ =\; & -\frac{2}{\Delta^{1/2}}\sum_{k=1}^{K}\frac{\sum_{i=2}^{M_\Delta} g_i^*\left\{(X_{(k-1)M_\Delta+i\Delta} - X_{(k-1)M_\Delta+(i-1)\Delta})^2 - \sigma^2\left(X_{(k-1)M_\Delta+(i-1)\Delta};\theta_2\right)\Delta\right\}}{\frac{r_k^2}{\Delta}}\end{aligned} \tag{36}$$

and, in view of Theorem 1 of BNS,

$$\frac{1}{\Delta^{1/2}}\sum_{i=2}^{M_\Delta} g_i^*\Delta\left\{(X_{(k-1)M_\Delta+i\Delta}-X_{(k-1)M_\Delta+(i-1)\Delta})^2-\sigma^2\left(X_{(k-1)M_\Delta+(i-1)\Delta};\theta_2\right)\Delta\right\}$$

$$\xrightarrow{d}\sqrt{2}\int_{(k-1)\frac{T}{K}}^{k\frac{T}{K}}\frac{\partial\sigma^2\left(X_s;\theta_2^*\right)}{\partial\theta_2}\sigma^2\left(X_s;\theta_2^*\right)dW_s \tag{37}$$

where W_s is a standard Brownian motion independent of X_t. It follows from (34), (37)and (36) that

$$\frac{1}{\Delta^{3/2}}\frac{\partial Q_\Delta\left(\theta_2^*\right)}{\partial\theta_2}\xrightarrow{d}\sum_{k=1}^{K}\frac{\sqrt{2}\int_{(k-1)\frac{T}{K}}^{k\frac{T}{K}}\frac{\partial\sigma^2\left(X_s;\theta_2^*\right)}{\partial\theta_2}\sigma^2\left(X_s;\theta_2^*\right)dW_s}{2\int_{(k-1)\frac{T}{K}}^{k\frac{T}{K}}\sigma^4\left(X_s;\theta_2^*\right)ds}. \tag{38}$$

Next we have

$$\begin{aligned}
\frac{1}{\Delta}\frac{\partial^2 Q_\Delta\left(\theta_2^*\right)}{\partial\theta_2\partial\theta_2'} &\sim \frac{2}{\Delta}\sum_{k=1}^{K}\frac{\sum_{i=2}^{M_\Delta}g_i^*g_i^{*\prime}\Delta^2}{\frac{r_k^2}{\Delta}}\\
&= 2\sum_{k=1}^{K}\frac{\sum_{i=2}^{M_\Delta}g_i^*g_i^{*\prime}\Delta}{\frac{r_k^2}{\Delta}}\\
&\xrightarrow{p} 2\sum_{k=1}^{K}\frac{\int_{(k-1)\frac{T}{K}}^{k\frac{T}{K}}\frac{\partial\sigma^2\left(X_s;\theta_2^*\right)}{\partial\theta_2}\frac{\partial\sigma^2\left(X_s;\theta_2^*\right)}{\partial\theta_2'}ds}{2\int_{(k-1)\frac{T}{K}}^{k\frac{T}{K}}\sigma^4\left(X_s;\theta_2^*\right)ds}\\
&= \sum_{k=1}^{K}\frac{\int_{(k-1)\frac{T}{K}}^{k\frac{T}{K}}\frac{\partial\sigma^2\left(X_s;\theta_2^*\right)}{\partial\theta_2}\frac{\partial\sigma^2\left(X_s;\theta_2^*\right)}{\partial\theta_2'}ds}{\int_{(k-1)\frac{T}{K}}^{k\frac{T}{K}}\sigma^4\left(X_s;\theta_2^*\right)ds}.
\end{aligned} \tag{39}$$

Combining (35), (38) and (39) we obtain

$$\Delta^{-1/2}\left(\hat{\theta}_2-\theta_2^*\right)\xrightarrow{d}\left[\sum_{k=1}^{K}\frac{\int_{(k-1)\frac{T}{K}}^{k\frac{T}{K}}\frac{\partial\sigma^2\left(X_s;\theta_2^*\right)}{\partial\theta_2}\frac{\partial\sigma^2\left(X_s;\theta_2^*\right)}{\partial\theta_2'}ds}{\int_{(k-1)\frac{T}{K}}^{k\frac{T}{K}}\sigma^4\left(X_s;\theta_2^*\right)ds}\right]^{-1}$$

$$\times\left[\sum_{k=1}^{K}\frac{\sqrt{2}\int_{(k-1)\frac{T}{K}}^{k\frac{T}{K}}\frac{\partial\sigma^2\left(X_s;\theta_2^*\right)}{\partial\theta_2}\sigma^2\left(X_s;\theta_2^*\right)dW_s}{2\int_{(k-1)\frac{T}{K}}^{k\frac{T}{K}}\sigma^4\left(X_s;\theta_2^*\right)ds}\right],$$

as stated. ∎

Proof of Theorem 4.4:

The proof follows similar lines to the proof of Theorem 4.2 and is therefore omitted. ∎

REFERENCES

Ahn, D. and B. Gao, 1999, A parametric nonlinear model of term structure dynamics. *Review of Financial Studies*, 12, 721–762.

Aït-Sahalia, Y., 1999, Transition densities for interest rate and other nonlinear diffusions. *Journal of Finance*, 54, 1361–1395.

Aït-Sahalia, Y., 2002, Maximum likelihood estimation of discretely sampled diffusion: A closed-form approximation approach. *Econometrica*, 70, 223–262.

Bandi, F. M. and P.C.B. Phillips, 2003, Fully nonparametric estimation of scalar diffusion models. *Econometrica*, 71, 241–283.

Bandi, F. M. and P.C.B. Phillips, 2004, A simple approach to the parametric estimation of potentially nonstationary diffusions. Unpublished paper, Cowles Foundation for Research in Economics, Yale University.

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

Barndorff-Nielsen, O. and N. Shephard, 2005, How accurate is the asymptotic approximation to the distribution of realised volatility?. In *Identification and Inference for Econometric Models*, eds by D.W.K. Andrews, J. Powell, P. Ruud and J. Stock, Cambridge University Press.

Billingsley, P., 1961, *Statistical Inference for Markov Processes*. Chicago, US: University of Chicago Press.

Bollerslev, T. and H. Zhou, 2002, Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *Journal of Econometrics*, 109, 33-65.

Brandt, M. W. and P. Santa-Clara, 2002, Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. *Journal of Financial Economics*, 30, 161–210.

Chan, K. C., G. A. Karolyi, F. A. Longstaff, and A. B. Sanders, 1992, An empirical comparison of alternative models of short term interest rates. *Journal of Finance*, 47, 1209–1227.

Cox, J., Ingersoll, J., and S. Ross, 1985, A theory of the term structure of interest rates. *Econometrica*, 53, 385–407.

Dixit, A., 1993, *The Art of Smooth Pasting*. Readings, UK: Harwood Academic Publishers.

Elerian, O., S. Chib, and N. Shephard, 2001, Likelihood inference for discretely observed non-Linear diffusions. *Econometrica*, 69, 959-993.

Eraker, B., 2001, MCMC analysis of diffusion models with application to finance. *Journal of Business and Economic Statistics*, 19, 177-191.

Florens-Zmirou, D., 1989, Approximate discrete-time schemes for statistics of diffusion processes. *Statistics*, 20, 547-557.

Hutton, J.E. and P. Nelson, 1986, Quasi-likelihood estimation for semimartingales. *Stochastic Processes and their Applications*, 22, 245-257.

Jacod, J. and P. Protter, 1998, Asymptotic error distributionbs for the Euler method for stochastic differential equations. *Annals of Probabilities*, 26, 267-307.

Lánska, V., 1979, Minimum contrast estimation in diffusion processes. *Journal of Applied Probability*, 16, 65–75.

Liptser, R.S. and A.N. Shiryaev, 2000, *Statistics of Random Processes*, Springer-Verlag, New York.

Lo, Andrew W., 1988, Maximum likelihood estimation of generalized Itô processes with discretely sampled data. *Econometric Theory*, 4, 231–247.

Merton, R. C. 1990, *Continuous-Time Finance*. Blackwell, Oxford.

Newey, W. K. and D. McFadden, 1994, Large sample estimation and hypothesis testing. In Engle, R.F. and D. McFadden, editors, *Handbook of Econometrics, Vol 4*. North-Holland, Amsterdam.

Pedersen, A., 1995, A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observation. *Scandinavian Journal of Statistics*, 22, 55-71.

Phillips, P.C.B. and J. Yu, 2004, Jackknifing bond option prices. *Review of Financial Studies*, forthcoming.

Prakasa Rao, B.L.S., 1999a, *Statistical Inference for Diffusion Type Processes*. Arnold, London.

Prakasa Rao, B.L.S., 1999b, *Semimartingales and Their Statistical Inference*. Chapman and Hall, Boca Raton, Florida

Vasicek, O., 1977, An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5, 177–186.

Yoshida, N., 1992, Estimation for diffusion processes from discrete observation. *Journal of Multivariate Analysis*, 41, 220–242.