Estimation of Longrun Variance of Continuous Time Stochastic Process Using Discrete Sample

Ye Lu†
Department of Economics
Indiana University

Joon Y. Park‡
Department of Economics
Indiana University
and Sungkyunkwan University

November 1, 2015

Abstract

This paper develops the methodology and asymptotic theory for the estimation of longrun variance of continuous time process. We analyze the asymptotic bias and variance of the longrun variance estimator in continuous time, and provide the optimal bandwidth balancing them off and minimizing the asymptotic mean squared error. In the paper, we present not only how to consistently estimate the longrun variance of continuous time process, but also how to choose bandwidth optimally with data dependent procedures, using discrete samples. Our framework is also useful to analyze the high frequency behaviors of usual longrun variance estimators for discrete time series. In particular, we show that they all diverge to infinity as the sampling frequency increases. The relevance and usefulness of our continuous time framework and asymptotic theory are demonstrated by illustration and simulation.

JEL Classification: C13, C22

Keywords: continuous time model, longrun variance estimator, kernel estimation, bandwidth selection


†Bloomington, IN 47405-7104. E-mail: lu33@indiana.edu.
‡Bloomington, IN 47405-7104. E-mail: joon@indiana.edu.
1. Introduction

The longrun variance (LRV) of a time series, measuring its total serial dependence, typically enters as a nuisance parameter into the limiting distribution of a statistic defined in a broad class of econometric models involving stationary and nonstationary time series. Its consistent estimation is therefore necessary for asymptotic inference. The LRV of a time series is commonly estimated by a weighted sum of its sample autocovariances, with the weight given by a kernel function and bandwidth parameter. Such a method is referred in econometrics to collectively as the heteroskedasticity and autocorrelation consistent (HAC) estimation of variance, highlighting that it robustifies inference in time series models by allowing for the presence of general dependence and heterogeneity of unknown form. The important early contributions on the HAC estimation include White (1984), Newey and West (1987, 1994), Andrews (1991), Robinson (1991) and Andrews and Monahan (1992), among many others.\(^1\) In particular, Andrews (1991) and Newey and West (1994) propose feasible and data dependent optimal bandwidth choices, which is of great importance in practical applications.

In this paper, we present the methodology and asymptotic theory for HAC estimation of LRV of continuous time process. Though continuous time model is used widely for an extensive range of empirical studies in economics and finance, the asymptotic theory for HAC estimation of its LRV is not available in the literature. Our paper fills this important gap. Our asymptotic theory is developed in parallel with that of HAC estimator of discrete time series in Andrews (1991). We analyze the asymptotic bias and variance of LRV estimator of continuous time process, and obtain the optimal bandwidth balancing them off and minimizing the asymptotic mean squared error. As usual, they are dependent on some unknown characteristics of underlying process, which we should consistently estimate to implement them in practical applications. We provide two approaches to obtain feasible optimal bandwidths for LRV estimation in continuous time, which correspond respectively to Andrews (1991) and Newey and West (1994) used commonly in estimating LRV of discrete time series.

Virtually in all practical applications, we do not observe any continuous time process in continuous time. Therefore, inference has to be made using discrete samples. In the paper, we assume that we have a discrete sample collected from a continuous time process at sampling interval \(\delta\) over time span \(T\), which yields a sample of size \(n = T/\delta\). Our

\(^1\)There have also been other approaches suggested in the literature, though we do not consider them in the paper. For example, Phillips (2005) proposed a LRV estimator as the explained sum of squares in a linear regression of the time series of interest on a set of regressors that are trend basis functions. See also Jansson (2002).
asymptotic theory is then developed for the LRV estimator obtained from this discrete sample as we let $\delta \to 0$ and $T \to \infty$.\(^2\) In this framework, we show that the estimator defined as $\Pi_{n,\delta} = \delta \Omega_n$, where $\Omega_n$ is the usual discrete time LRV estimator using bandwidth $b_n = B_T/\delta$ with $B_T$ such that $B_T \to \infty$ and $B_T/T \to 0$ as $T \to \infty$, is generally consistent for the LRV $\Pi$ of continuous time process. Consistent estimation for the LRV of continuous time process using its discrete sample is therefore simple and straightforward. It is also easy to obtain feasible and data dependent optimal bandwidths using a discrete sample, which yield a LRV estimator having the smallest asymptotic mean squared error.

As a byproduct, our continuous time framework provides a useful tool to analyze the frequency dependence of the LRV estimates of discrete time series. The LRV estimates of many economic and financial time series are highly sensitive to the sampling frequency. In particular, they tend to increase and diverge to infinity as the sampling interval $\delta$ decreases and approaches to 0. This is entirely consistent with our asymptotic theory based on a continuous time framework. Note that the consistency of $\Pi_{n,\delta} = \delta \Omega_n$ for the LRV $\Pi$ of underlying continuous time process implies that the LRV estimate $\Omega_n$ of its discrete sample diverges to infinity as $\delta \to 0$ and does not consistently estimate any well defined parameter. The frequency dependence of the discrete time LRV estimates of actual economic and financial time series disappears instantly, if we regard them as discrete samples from some underlying continuous time processes and obtain the continuous time LRV estimates of underlying processes. They are indeed very stable and invariant across a wide range of different sampling frequencies.

To show the practical relevance and usefulness of our continuous time framework and asymptotic theory, we analyze two leading financial time series and simulate two commonly used diffusion models.\(^3\) We observe that the discrete time LRV estimates of both financial time series exhibit clear upward trends as the sampling frequency increases, though their continuous time LRV estimates are rather stable across different sampling frequencies. This feature is well replicated in our simulation. We also study through our simulation the finite sample performances of the LRV estimators of continuous time process relying on different models and bandwidth choices for various combinations of $\delta$ and $T$. Our simulation results are entirely consistent with our asymptotic theory except for the case of $T$ being very small. The effect of $\delta$ is minimal, unless the underlying process has excessively strong mean reversion. In all other cases where the underlying process has realistic mean reversion, the

\(^2\)Our asymptotics are joint, not sequential, in the sense that we let $\delta \to 0$ and $T \to \infty$ simultaneously. We only require $\delta \to 0$ fast enough relative to $T \to \infty$.

\(^3\)The time series we analyze are 1-month forward premium of US/UK exchange rates and 3-month US T-bill rates. We also use them to fit Ornstein-Uhlenbeck and Feller’s Square Root processes and get the parameter values of the baseline models in our simulations.
LRV estimators obtained from quarterly observations perform well enough and have roughly
equally mean squared errors to those obtained from monthly or even daily observations.

The rest of the paper is organized as follows. Section 2 introduces the LRV of continuous
time process and its HAC estimator, and analyzes the asymptotic bias and variance of HAC
estimator. Subsequently, Section 3 presents the optimal bandwidth, which balances off the
asymptotic bias and variance of HAC estimator and yields the HAC estimator minimizing
the asymptotic mean squared error. Discussions on how to obtain feasible and data de-
dendent optimal bandwidths are also provided. Section 4 considers the LRV estimation of
continuous time process using discrete sample. In particular, we propose a LRV estimator of
continuous time process based on a discrete sample from the underlying process, and estab-
lish its consistency under very general conditions. In Section 5, we summarize and report
our empirical illustrations and simulation results. Section 6 concludes the paper. Finally,
all mathematical proofs and additional figures are collected and presented in Appendix.

2. Estimation of Longrun Variance in Continuous Time

We let \( U \) be a mean zero \( d \)-dimensional continuous time process, and define a process
\( U^T = (U^T_t) \) on the unit interval \([0, 1] \) as

\[
U^T_t = \frac{1}{\sqrt{T}} \int_0^t U_s ds,
\]

(1)

where \( U^T \) is regarded as a random element taking values in \( D^d[0, 1] \), the space of all \( d \)-
dimensional cadlag functions on the unit interval \([0, 1] \). Under suitable conditions, we have

\[
U^T \rightarrow_d U^o
\]

(2)
as \( T \rightarrow \infty \), due to the continuous time version of Donsker’s invariance principle, where \( U^o \)
is a Brownian motion on \([0, 1] \) with covariance matrix

\[
\Pi = \lim_{T \rightarrow \infty} \frac{1}{T} E \left\{ \left( \int_0^T U_t dt \right) \left( \int_0^T U_t dt \right)^T \right\},
\]

(3)

which is assumed to exist. The covariance matrix \( \Pi \) of limit Brownian motion \( U^o \) in (3) is
often called the longrun variance (LRV) of underlying process \( U \), which we need to estimate
for statistical inference on various aspects of models involving \( U \).

There has been a large literature on the invariance principle introduced in (2). In par-
ticular, it is well established for a wide class of processes with verifiable mixing conditions,
see, e.g., a survey by Merlevède et al. (2006)). For instance, the invariance principle holds for a mean zero stationary process with finite second moment and $\rho(t)$ satisfying the condition $\int_0^\infty \rho(2^t)dt < \infty$, which is clearly satisfied for any $\rho(t)$ with a polynomial decay rate $\rho(t) = O(t^{-p})$ for any $p > 0$. Also, for a mean zero and strong mixing process, the invariance principle holds if it has bounded $2p$-th moment, and its strong mixing coefficients $\alpha(t)$ satisfy the condition $\int_1^\infty \alpha(t)^{1-1/p}dt < 1$ with $p > 1$, which allows for the polynomial decay rate given by $\alpha(t) = O(t^{-q})$ for some $q > p/(p-1)$ with $p > 1$.

The invariance principle for the class of functions of continuous time stationary ergodic Markov process is provided by Bhattacharya (1982). For a continuous time stationary ergodic Markov process $X$, which has state space $\mathcal{D}$ with speed measure $m$ and infinitesimal generator $A$ defined on $L^2(\mathcal{D},dm)$, the invariance principle holds for $U = f(X)$ with any function $f$ defined on the range of $A$ such that $m(f) = 0$. Specifically, if $X$ is a mean zero positive recurrent diffusion process on $\mathcal{D}$ with drift function $\mu$ and diffusion function $\sigma$, for which the scale function $s$ and the speed density $m$ are given respectively by $s(x) = \int x \exp(-\int y(2\mu(z)/\sigma^2(z))dz)dy$ and $m(x) = 1/(\sigma^2 s'(x))$, the invariance principle holds for any process defined as $U = f(X)$ with $m(f) = 0$, provided its LRV

$$\Pi = 4m(\mathcal{D})^{-1} \int_\mathcal{D} \left( \int_0^T f(y)m(dy) \right)^2 s'(x)dx$$

is finite. See also e.g., (1.1) in van der Vaart and van Zanten (2005) and Corollary 3.6(a) in Kim and Park (2014).

Now we let the invariance principle (2) hold for a mean zero continuous time process $U$, and consider estimation of its LRV $\Pi$ in (3). Also, to focus on the development of required statistical theory, we tentatively assume that the sample path of $U$ is observed continuously in time. Notice that the estimand $\Pi$ can be written as

$$\Pi = \lim_{T \to \infty} \int_{-T}^T \mathbb{E}(\Gamma_T(s))ds,$$

where

$$\Gamma_T(s) = \begin{cases} 
\frac{1}{T} \int_{-T}^{-|s|} U_t U'_{t-s} dt, & \text{for } s \leq 0, \\
\frac{1}{T} \int_{s}^{T} U_t U'_{t-s} dt, & \text{for } s > 0,
\end{cases}$$

is the sample autocovariance function of $U$. When $U$ is second-order stationary, we have $\mathbb{E}(\Gamma_T(s)) = (1-|s|/T)\mathbb{E}(U_t U'_{t-s}) \to \Gamma(s) = \mathbb{E}(U_t U'_{t-s})$ as $T \to \infty$ pointwise in $s$. Therefore,

\footnote{In particular, if the Markov process $X$ is $\rho$-mixing, it can be shown that for all $f \in L^2(\mathcal{D},dm)$ such that $m(f) = 0$ the invariance principle holds for $U = f(X)$.}
it follows from the bounded convergence that \( \Pi = \int_{-\infty}^{\infty} \Gamma(s) ds \), which corresponds to the spectral density of \( U \) evaluated at frequency zero multiplied by \( 2\pi \). This motivates the use of a spectral density estimator to estimate \( \Pi \). In the paper, we consider the estimator \( \Pi_T \) for LRV \( \Pi \) of \( U \), which is given by

\[
\Pi_T = \int_{|s| \leq T} K \left( \frac{s}{B_T} \right) \Gamma_T(s) ds, \tag{5}
\]

where \( \Gamma_T(s) \) is the sample autocovariance function defined in (4) above, \( K \) denotes a kernel function, and \( B_T \) is a bandwidth parameter. Clearly, \( \Pi_T \) is the continuous time analogue of the discrete time HAC estimator considered in, e.g., Andrews (1991).

The bandwidth \( B_T \) serves as a trimming parameter for kernels that assign zero weights to values greater than 1.\(^5\) As in discrete time case, the choice of bandwidth \( B_T \) is a critical issue in continuous time LRV estimation. In particular, we have different asymptotics depending upon how we set \( B_T \) in relation to \( T \). If \( B_T/T \to 0 \) as \( T \to \infty \), we have small-\( b \) or increasing smoothing asymptotics. This is a conventional assumption to achieve consistency of LRV estimator \( \Pi_T \). On the other hand, we have fixed-\( b \) or fixed smoothing asymptotics in case that \( B_T/T \) does not tend to zero as \( T \) goes to infinity. Under fixed-\( b \) scheme, LRV estimator is no longer consistent. However, this scheme has also been used widely especially in developing the asymptotic theory for hypothesis testing, since it yields more relevant asymptotics than small-\( b \) scheme by taking into account the finite sample variation of LRV estimator used in a test.\(^6\) Therefore, we provide both asymptotics in the paper.

For a kernel function \( K: \mathbb{R} \to [-1,1] \), we let

\[
\pi(s) = \lim_{x \to 0} \frac{1 - K(x)}{|x|^s}
\]

for nonnegative integer \( s \), and define the characteristic exponent \( r \) of \( K \) by \( r = \max\{s : \pi(s) < \infty\} \). If \( \pi(s) = \infty \) for any nonnegative integer \( s \), we set \( r = \infty \). The characteristic exponent \( r \) is a smoothness measure of a kernel function around zero. For the development of our asymptotics in the paper, we need to introduce some technical assumptions.

**Assumption 2.1.** \( K \) is symmetric with \( K(0) = 1 \) and \( \int K^2(x) dx < \infty \), and continuous at 0 and at all but a finite number of other points.

\(^5\)There exist kernels such as quadratic spectral kernel that are nonzero over the entire real line.

\(^6\)For more discussions on the usefulness of fixed-\( b \) asymptotics, the reader is referred to Kiefer and Vogelsang (2002, 2005), Phillips et al. (2006, 2007), Sun et al. (2008, 2011) and Sun (2014), among many others.
Assumption 2.2. We assume
(a) \( K \) has a finite characteristic exponent \( r > 0 \), and
(b) \( \int_{-\infty}^{\infty} |s|^r \sup_{t \geq 0} \| \mathbb{E} U_t U'_{t-s} \| \, ds < \infty \) for the characteristic exponent \( r > 0 \) of \( K \).

The conditions in Assumptions 2.1 and 2.2 are mild and satisfied widely. Assumption 2.1 holds for many commonly used kernels such as truncated, Bartlett, Parzen and quadratic spectral kernels. Moreover, Part (a) of Assumption 2.2 is satisfied by virtually all kernels used in practical applications except for truncated kernel that has \( r = 1 \). We have \( r = 1 \) for Bartlett kernel with \( \pi(1) = 1 \), and \( r = 2 \) for Parzen kernel with \( \pi(2) = 6 \), Tukey-Hanning kernel with \( \pi(2) = \pi^2/4 \), and quadratic spectral kernel with \( \pi(2) \approx 1.4212 \). The characteristic exponent \( r \) determines the convergence rate for the asymptotic bias of LRV estimator \( \Pi_T \) in (5). Part (b) of Assumption 2.2 requires that the autocovariances of underlying process \( U \) decay sufficiently fast to facilitate the asymptotic analysis for the bias of \( \Pi_T \).

The following lemma provides the asymptotic bias of \( \Pi_T \) under both small-\( b \) and fixed-\( b \) schemes.

Lemma 2.1. Let Assumptions 2.1 and 2.2 hold. We have

\[
\limsup_{T \to \infty} \frac{B_T}{T} \left[ \mathbb{E} \mathbb{v}'(\Pi_T - \Pi) \mathbb{v} \right] \leq \pi(r) \int_{-\infty}^{\infty} |s|^r \sup_{t \geq 0} \| \mathbb{E} v U_t U'_{t-s} \| \, ds
\]

for any \( v \in \mathbb{R}^d \), under small-\( b \) asymptotics with \( B_T/T \to 0 \) and under fixed-\( b \) asymptotics with \( B_T/T = b > 0 \), respectively.

If \( U \) is stationary, we have

\[
\lim_{T \to \infty} \frac{B_T}{T} \left( \mathbb{E} \Pi_T - \Pi \right) = -\pi(r) \int_{-\infty}^{\infty} |s|^r \Gamma(s) \, ds
\]

respectively under small-\( b \) and fixed-\( b \) asymptotics. The asymptotics for the bias of LRV estimator are given in Parzen (1957) under small-\( b \) scheme, and in Neave (1970) under fixed-\( b \) scheme.

Lemma 2.1 implies, in particular, that under small-\( b \) asymptotics the magnitude of bias in \( \Pi_T \) diminishes faster if the characteristic exponent \( r \) of \( K \) increases and \( K \) has a smoother curvature around the origin. As shown in the proof of Lemma 2.1, the bias of \( \Pi_T \) arises
mainly from the sub-unit weights assigned to the included sample autocovariances $\Gamma_T(s)$ for $|s|$ smaller than $T$, which are of orders $O(B_T^{-r})$ and $O(B_T^{-1})$ respectively for small-$b$ and fixed-$b$ asymptotics. There are two other bias terms: one due to trimming all sample autocovariances $\Gamma_T(s)$ for $|s|$ bigger than $T$,\(^7\) and the other incurred by using $T$ instead of $T - |s|$ in defining $\Gamma_T(s)$ for all $s > 0$. Both of them are of orders less than the leading bias term and asymptotically negligible.

We need the following assumption to obtain the asymptotic variance of $\Pi_T$.

**Assumption 2.3.** We assume
\begin{enumerate}[(a)]  
  \item $\int_{-\infty}^{\infty} \sup_{t \geq 0} \left\| U_t U'_{t-s} \right\| ds < \infty$, and  
  \item $\int_{-\infty}^{\infty} \sup_{t \geq 0} \left| C_{ijhk}(t, t+s_1, t+s_2, t+s_3) - C_{ijhk}^*(t, t+s_1, t+s_2, t+s_3) \right| ds_1 ds_2 ds_3 < \infty$, where $C_{ijhk}(t, t+s_1, t+s_2, t+s_3)$ and $C_{ijhk}^*(t, t+s_1, t+s_2, t+s_3)$ are the fourth-order moments of $(U_{i,t}, U_{j,t+s_1}, U_{h,t+s_2}, U_{k,t+s_3})$ and $(U_{i,t}^*, U_{j,t+s_1}^*, U_{h,t+s_2}^*, U_{k,t+s_3}^*)$ for all $i, j, h, k = 1, \ldots, d$, defined with a Gaussian process $U^* = (U_t^*)$ with the same mean and covariance functions as $U = (U_t)$ for $i = 1, \ldots, m$.
\end{enumerate}

We may readily extend Lemma 1 in Andrews (1991) to continuous time models, and show that Assumption 2.3 holds for a continuous time process $U$, if its continuous part is a strong mixing process such that, for some $p > 1$, it has bounded $4p$-th moment and its strong mixing coefficients satisfy the condition $\int_0^\infty t^2 \alpha(t)^{(p-1)/p} < \infty$, and its jump part is given by an independent Lévy process such that $\sum_{0 \leq t \leq T} E \left\| U_t \right\|^{4p} = O(T).\(^8\)$ These conditions are not stringent and expected to hold for a wide class of continuous time processes. It is natural to require that $U$ has bounded moments up to at least the fourth order to analyze the variance of $\Pi_T$. Of course, Part (b) is trivially satisfied for all continuous time Gaussian processes.

The asymptotic variance of LRV estimator $\Pi_T$ is provided in the following lemma.

**Lemma 2.2.** *Let Assumptions 2.1 and 2.3 hold. We have*
\begin{align*}
\limsup_{T \to \infty} \frac{T}{B_T} \text{var } (v' \Pi_T v) &\leq 2 \left( \int_{-\infty}^{\infty} \sup_{t \geq 0} \left| E v' U_t U'_{t-s} v \right| ds \right)^2 \int_{-\infty}^{\infty} K^2(x) dx \\
\limsup_{T \to \infty} \text{var } (v' \Pi_T v) &\leq 2b \left( \int_{-\infty}^{\infty} \sup_{t \geq 0} \left| E v' U_t U'_{t-s} v \right| ds \right)^2 \int_{-1/b}^{1/b} K^2(x) dx
\end{align*}

\(^7\)For the kernels that assign zero to all the values greater than 1, all the sample autocovariances of orders larger than $B_T$ are trimmed.  
\(^8\)Note that the presence of Lévy jumps in the continuous time process will only affect the instantaneous variance, but not the serial dependence of the process. Moreover, all processes with compound Poisson type jumps whose jump sizes have finite $4p$-th moments and jump intensity proportional to $T$ satisfy the moment condition here.
for any $v \in \mathbb{R}^d$, under small-$b$ asymptotics with $B_T^r/T \to 0$ and under fixed-$b$ asymptotics with $B_T/T = b > 0$, respectively.

If $U$ is stationary, we have

$$
\lim_{T \to \infty} \frac{T}{B_T} \text{var}(\text{vec} \Pi_T) = (I + K_d)(\Pi \otimes \Pi) \int_{-\infty}^{\infty} K^2(x)dx
$$

$$
\lim_{T \to \infty} \text{var}(\text{vec} \Pi_T) = b(I + K_d)(\Pi \otimes \Pi) \int_{-1/b}^{1/b} (1 - b|x|)K^2(x)dx,
$$

respectively under small-$b$ and fixed-$b$ asymptotics, where $K_d$ denotes commutation matrix of dimension $d^2 \times d^2$. The asymptotic variances of LRV estimator under small-$b$ and fixed-$b$ schemes are provided respectively in Parzen (1957) and Neave (1970).

Lemma 2.2 shows that the asymptotic variance of $\Pi_T$ is of order $O(B_T/T)$ and it decreases down to zero in the limit under small-$b$ asymptotics, since bandwidth parameter $B_T$ increases at a slower rate than time span $T$. However, under fixed-$b$ asymptotics, where bandwidth parameter $B_T$ grows at the same rate as time span $T$, the asymptotic variance of $\Pi_T$ does not vanish and $\Pi_T$ remains to be random in the limit. See Andrews (1991) and Neave (1970) for the corresponding results obtained in the usual discrete time framework.

Due to Lemmas 2.1 and 2.2, it is straightforward to establish consistency of $\Pi_T$ under small-$b$ asymptotics for a general continuous time process possibly with some moderate nonstationarity. In what follows, we let $U^0$ be defined as in (2).

**Theorem 2.3.** Let Assumptions 2.1, 2.2 and 2.3 hold. Under small-$b$ asymptotics with bandwidth $B_T$ satisfying $B_T \to \infty$ and $B_T^r/T \to 0$ as $T \to \infty$, we have $\Pi_T \to_p \Pi$ as $T \to \infty$. On the other hand, under fixed-$b$ asymptotics with bandwidth $B_T$ satisfying $B_T/T = b > 0$, we have $\Pi_T \to_d \int_0^1 \int_0^1 K((t - s)/b)dW_t^\alpha dW_s^\alpha$ as $T \to \infty$.

Our result under small-$b$ asymptotics in Theorem 2.3 is entirely analogous to the corresponding result for discrete time LRV estimator in Andrews (1991).\footnote{We require $B_T^r/T \to 0$ as $T \to \infty$ for our small-$b$ asymptotics here to deduce the consistency of $\Pi_T$ from Lemmas 2.1 and 2.2. It can, however, be established under a weaker condition $B_T/T \to 0$ as $T \to \infty$ following Hansen (1992) and de Jong and Davidson (2000).} The result under fixed-$b$ asymptotics is obtained in the spirit of Kiefer and Vogelsang (2005). Note that we may write $U^0 = \Pi^{1/2}W$ with $d$-dimensional standard Brownian motion $W$ to have $\Pi_T \to_d \Pi^{1/2} \int_0^1 \int_0^1 K((t - s)/b)dW_t^\alpha dW_s^\alpha \Pi^{1/2}$. As expected, $\Pi_T$ is not consistent under fixed-$b$ asymptotics.\footnote{It is not difficult to show that $\int_0^1 \int_0^1 K((t - s)/b)dW_t dW_s' \to_p 1$ as $b \to 0$. Of course, in this case, fixed-$b$ asymptotics reduce to small-$b$ asymptotics and we have $\Pi_T \to_p \Pi$.}
3. Optimal Bandwidth and Bandwidth Selection

In this section, we introduce the asymptotic theory of optimal bandwidth and discuss how to choose bandwidth optimally in practice. Throughout this section, we assume that $U$ is stationary and has autocovariance function $\Gamma$, and define

$$\Lambda(r) = \int_{-\infty}^{\infty} |s|^r \Gamma(s) ds.$$  \hfill (6)

Moreover, for a matrix $A$, we let its norm be given by $\|A\|_M = \text{tr}(A'MA)^{1/2}$ with some fixed positive definite matrix $M$, so that the mean squared error (MSE) of $\hat{\tau}$ is given by

$$\text{MSE}(\hat{\tau}, M) = \mathbb{E}\|\hat{\tau} - \tau\|^2_M = \mathbb{E} \left[ \text{tr}(\hat{\tau}^T - \tau) M (\hat{\tau}^T - \tau) \right]$$

accordingly.

If we set $B_T^{2r+1}/T \to \lambda > 0$ as $T \to \infty$ to balance off the bias and variance terms derived respectively in Lemmas 2.1 and 2.2, it follows that

$$\lim_{T \to \infty} T^{2r/(2r+1)} \text{MSE}(\hat{\tau}, M) = \lambda^{1/(2r+1)} \text{tr}(I \otimes M)(I + K_d)(\Pi \otimes \Pi) \int_{-\infty}^{\infty} K^2(x) dx$$

$$+ \pi^2(r)(\text{vec} \Lambda(r))' (I \otimes M)(\text{vec} \Lambda(r))/\lambda^{2r/(2r+1)},$$

which is minimized over $\lambda$ at

$$\lambda^* = \frac{r \pi^2(r)}{\int_{-\infty}^{\infty} K^2(x) dx} \Theta^2(r),$$  \hfill (7)

where

$$\Theta^2(r) = \frac{2(\text{vec} \Lambda(r))' (I \otimes M)(\text{vec} \Lambda(r))}{\text{tr}(I \otimes M)(I + K_d)(\Pi \otimes \Pi)}.$$  \hfill (8)

Therefore, given a kernel function $K$ and a positive definite weighting matrix $M$, the optimal bandwidth $B_T^*$ that minimizes the asymptotic MSE of $\Pi_T$ can be obtained as

$$B_T^* = (\lambda^* T)^{1/(2r+1)}$$  \hfill (9)

with $\lambda^*$ defined in (7).

Notice that the growth rate of optimal bandwidth depends on the characteristic exponent $r$ of the kernel function $K$ used. In fact, the smoother the kernel function is around zero, the faster the optimal bandwidth shrinks relative to the total time span. The constant $\lambda^*$ is given in (8) as the product of two terms: $r \pi^2(r)/\int_{-\infty}^{\infty} K^2(x) dx$ and $\Theta^2(r)$. The first
term is fully characterized by the kernel function, which is given by 1.50, 133.51, 16.23 and 4.04 respectively for Bartlett, Parzen, Tukey-Hanning and quadratic spectral kernels. On the other hand, the second term \( \Theta^2(r) \) is determined by LRV \( \Pi \) and \( \Lambda(r) \) defined in (6).

Clearly, \( B_T^* \) defined in (9) is infeasible, since \( \Theta^2(r) \) depends on the autocovariance structure and LRV of underlying process \( U \) and therefore is generally unknown. To make it feasible, we introduce a data dependent procedure for the choice of bandwidth similarly as in Andrews (1991) and Newey and West (1994). It is straightforward to extend the approach by Newey and West (1994) to our continuous time framework. If continuous time observations are available, we may estimate \( \lambda^* \) in (6) by

\[
\lambda^*_T(r) = \sum_{i=1}^{d} \int_{-\infty}^{\infty} \frac{|s|^{\Gamma_i(s)}}{\Gamma_{\infty}} ds,
\]

(10)

where \( \Gamma_T \) is the sample autocovariance function of \( U \) in continuous time, and \( A_T \) is set such that \( A_T \to \infty \) and \( A_T/T \to 0 \) as \( T \to \infty \). Under appropriate regularity conditions, we would expect \( \Lambda_T(r) \to \Lambda(r) \) as \( T \to \infty \), and therefore, we may use the feasible optimal bandwidth \( \tilde{B}_T^* = (\lambda^*_T)^{1/(2r+1)} \), where \( \lambda^*_T \) is defined as in \( \lambda^* \) in (6) with \( \Lambda(r) \) replaced by \( \Lambda_T(r) \) in (10). In our subsequent discussions, \( \tilde{B}_T^* \) will be referred to as the continuous time nonparametric (CNP) bandwidth.

The continuous time extension of the bandwidth selection procedure by Andrews (1991) is more involved. For the required extension, we set \( M = \text{diag}(w_1, \ldots, w_d) \), so that \( \Theta^2(r) \) reduces to

\[
\Theta^2(r) = \sum_{i=1}^{d} w_i \left( \int_{-\infty}^{\infty} |s|^{\Gamma_i(s)} ds \right)^2 / \left( \sum_{i=1}^{d} w_i \left( \int_{-\infty}^{\infty} \Gamma_i(s) ds \right)^2 \right).
\]

(11)

where \( \Gamma_i \) denotes the autocovariance function of \( U_i \), i.e., the \( i \)-th component of \( U \), \( i = 1, \ldots, d \).

Momentarily, we let \( U \) be one-dimensional and assume that it is generated as a jump diffusion process given by

\[
dU_t = -\kappa U_t dt + \tau(U_t) dW_t + v(U_t) dJ_t,
\]

(12)

where \( \kappa > 0 \) is mean reversion parameter, \( \tau \) and \( v \) signify respectively diffusive and jump volatility functions, \( W \) is standard Brownian motion and \( J \) is a general Lévy jump process further specified as

\[
dJ_t = \int_{\mathbb{R}} x N(dt, dx)
\]

with Poisson random measure \( N \) that has associated Lévy measure \( \nu \), i.e., \( \nu(dx) dt = \)
\( \mathbb{E} N(dt, dx) \), satisfying the conditions \( \int_{\mathbb{R}} x \nu(dx) = 0 \) and \( \int_{\mathbb{R}} x^2 \nu(dx) = 1 \). As a special case, we may consider \( J \) defined more explicitly as \( dJ_t = Z_t dN_t \), where \( Z \) is an i.i.d. stochastic process distributed as \( \nu/\nu(\mathbb{R}) \) and \( N \) is Poisson process with arrival rate \( \nu(\mathbb{R}) \). Our jump diffusion model is quite general, allowing for arbitrary forms of state dependence for both diffusive and jump volatilities, except that it is required to have a linear drift. Throughout, we assume that \( U \) in (12) is stationary and has a finite variance \( \sigma^2 > 0 \).

**Lemma 3.1.** Let \( U \) be generated as in (12). Then we have

\[
\Gamma(s) = \sigma^2 e^{-\kappa |s|}
\]

for \(-\infty < s < \infty\).

As shown in Lemma 3.1, the autocovariance function of \( U \) introduced in (12) is therefore dependent only upon its mean reversion parameter \( \kappa > 0 \) and variance \( \sigma^2 \). It follows immediately from Lemma 3.1 that

\[
\Pi = \int_{-\infty}^{\infty} \Gamma(s)ds = \frac{2\sigma^2}{\kappa} \quad \text{and} \quad \Lambda(r) = \int_{-\infty}^{\infty} |s|^r \Gamma(s)ds = \frac{2r!\sigma^2}{\kappa^{r+1}} \tag{13}
\]

for any positive integer \( r \). Therefore, we may readily deduce that

**Proposition 3.2.** We have

\[
\Theta^2(r) = \frac{\sum_{i=1}^{d} w_i (r!)^2 \left( \frac{\sigma_i^4}{\kappa_i^2} \right)^{(r+1)}}{\sum_{i=1}^{d} w_i \left( \frac{\sigma_i^4}{\kappa_i^2} \right)}
\]

if \( \Theta^2(r) \) is defined as in (11) for any positive integer \( r > 0 \), and if each component \( U_i \) of \( U = (U_i) \) is generated as in (12) with mean reversion parameter \( \kappa_i \) and variance \( \sigma_i^2 \) for \( i = 1, \ldots, d \).

In light of Proposition 3.2, the automatic feasible optimal bandwidth \( \tilde{B}_T^* \) is now easy to obtain if each component \( U_i \) of \( U = (U_i) \) is generated as in (12). We may first estimate \( \kappa_i \) and \( \sigma_i^2 \) for each component \( U_i \) of \( U \), and then plug in their consistent estimates \( \kappa_{iT} \) and \( \sigma_{iT}^2 \) defined as

\[
\kappa_{iT} = -\left( \int_0^T U_{it}^2 dt \right)^{-1} \int_0^T U_{it} dU_{it} \quad \text{and} \quad \sigma_{iT}^2 = \frac{1}{T} \int_0^T U_{it}^2 dt,
\]

\footnote{For Poisson random measure \( N \), \( N((0,t], A) \) represents the number of jumps with sizes in \( A \) on the interval \((0,t] \) for any \( t \in \mathbb{R}_+ \) and \( A \in \mathcal{B}(\mathbb{R}) \), and Lévy measure \( \nu \) associated with \( N \) is defined as \( \nu(A) = \mathbb{E} N((0,1], A) \) for any \( A \in \mathcal{B}(\mathbb{R}) \).}
Table 1. Optimal bandwidth $B_T^*$ for LRV estimate of univariate jump diffusion having linear drift with mean reversion parameter $\kappa$ using Bartlett, Parzen and QS kernels with sample span $T$.

<table>
<thead>
<tr>
<th>$T$</th>
<th>Bartlett Kernel</th>
<th>Parzen Kernel</th>
<th>QS Kernel</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>$\kappa = 0.5$</td>
<td>3.107 1.957 0.669</td>
<td>$\kappa = 0.5$</td>
</tr>
<tr>
<td>10</td>
<td>3.915 2.466 0.843</td>
<td>9.691 5.566 1.536</td>
<td>4.814 2.765 0.763</td>
</tr>
<tr>
<td>20</td>
<td>4.932 3.107 1.063</td>
<td>11.132 6.393 1.764</td>
<td>5.530 3.176 0.876</td>
</tr>
</tbody>
</table>

$i = 1, \ldots, d$, into our formula for $\Theta^2(r)$ to get $\tilde{\Theta}^2(r)$, respectively for $r = 1$ and 2. In what follows, we call $\tilde{B}_T^*$ with $\tilde{\Theta}^2(r)$ using $\kappa_{iT}$ and $\sigma^2_{iT}$ as estimates for $\kappa_i$ and $\sigma_i^2$, $i = 1, \ldots, d$, the continuous time semiparametric (CSP) bandwidth.

Proposition 3.2 implies that the optimal bandwidth $B_T^*$ for $\Pi_T$ depends only on the mean reversion parameter $\kappa$ of underlying process $U$, if $U$ is one-dimensional and generated as a stationary jump diffusion with linear drift as in (12). In fact, in this case, we have

$$\Theta^2(r) = (r!)^2/\kappa^{2r}$$

for any positive integer $r > 0$, and therefore, mean reversion parameter $\kappa$ alone determines the optimal bandwidth parameter.

For an illustration, Table 1 tabulates the asymptotically optimal bandwidth parameter $B_T^*$ for three most frequently used kernels – Bartlett, Parzen and QS kernels – when underlying process $U$ is specified as a stationary jump diffusion with linear drift and mean reversion parameter $\kappa$ as in (12). It can be seen that optimal bandwidth $B_T^*$ increases as sample span $T$ gets larger. More importantly, a smaller $\kappa$ implies higher level of persistence, and therefore a slower decay of temporal dependence, in underlying process $U$. Consequently, for a smaller $\kappa$, optimal bandwidth $B_T^*$ would be greater and more higher order terms of autocovariance functions should be included in estimating the LRV of $U$.

4. Longrun Variance Estimation Using Discrete Sample

All our previous discussions assume that underlying continuous time process $U$ is observed continuously in time. However, in virtually all practical applications, observations are made only in discrete time. In this section, we explain how we estimate LRV of continuous time process $U$ using observations collected from $U$ in discrete time intervals.
In our subsequent analysis, we let \((u_i)\) be a sample of size \(n\) collected at discrete time intervals from underlying continuous time process \(U\), i.e.,

\[
u_i = U_{i\delta},
\]

over time span \([0, T]\) with sampling interval \(\delta > 0\), where \(T = n\delta\), and define the LRV estimator of \((u_i)\) as

\[
\Omega_n = \sum_{|j| \leq n-1} K\left(\frac{j}{b_n}\right) \Gamma_n(j),
\]

where \(K\) is a kernel function, \(\Gamma_n(j) = (1/n) \sum_i u_i u'_{i-j}\) are sample autocovariances of \((u_i)\), and \(b_n\) is the bandwidth parameter defined as

\[
b_n = B_T/\delta
\]

from a given continuous time bandwidth parameter \(B_T\) such that \(B_T \to \infty\) and \(B_T/T \to 0\).

Note that we have

\[
\Gamma_n(j) = \frac{1}{n} \sum_{i=1}^{n} u_i u'_{i-j} = \frac{1}{T} \sum_{i=1}^{n} \delta U_{i\delta} U'_{(i-j)\delta} \approx \frac{1}{T} \int_0^T U_t U'_{t-j\delta} dt = \Gamma_T(j\delta)
\]

for \(\delta\) sufficiently small relative to \(T\). Therefore, if we define

\[
\Pi_{n,\delta} = \delta \Omega_n,
\]

it follows from (16) and (17) that

\[
\Pi_{n,\delta} \approx \sum_{|j| \leq n-1} \delta K\left(\frac{j\delta}{B_T}\right) \Gamma_T(j\delta) \approx \int_{|s| \leq T} K\left(\frac{s}{B_T}\right) \Gamma_T(s) ds.
\]

Consequently, we may use \(\Pi_{n,\delta}\) in (18) to estimate the LRV \(\Pi\) of continuous time process \(U\). We will show subsequently that, under suitable regularity conditions, the proposed estimator \(\Pi_{n,\delta}\) based on the discrete sample \((u_i)\) is indeed consistent for the LRV of underlying continuous time process \(U\).

To more formally present the asymptotic theory for the discrete sample LRV estimator \(\Pi_{n,\delta}\) of continuous time process, we introduce some technical conditions. Subsequently, we let

\[
\sup_{0 \leq s, t \leq T} \sup_{|t-s| \leq \delta} \mathbb{E} \|U_t - U_s\|^2 = O\left(\frac{\Delta^2_{\delta,T}}{T}\right),
\]

as \(\delta \to 0\) and \(T \to \infty\), where \(\Delta^2_{\delta,T}\) may be interpreted as the asymptotic modulus of
continuity of $U$ in expectation, as opposed to the usual modulus of continuity of a stochastic process for which $T$ is treated as fixed. We expect that $\Delta_{\delta,T}$ decreases as $\delta \to 0$ and increases as $T \to \infty$.

**Assumption 4.1.** $\sup_{t \geq 0} \mathbb{E} \|U_t\|^2 < \infty$.

**Assumption 4.2.** $K$ is differentiable with absolutely integrable derivative.

**Assumption 4.3.** $\Delta_{\delta,T} B_T \to 0$ as $\delta \to 0$ and $T \to \infty$.

The conditions in Assumptions 4.1, 4.2 and 4.3 are not stringent and expected to be satisfied for a broad class of continuous time stochastic processes. Assumption 4.1 is extremely mild and satisfied widely. It only requires the second moment of $U$ to be uniformly bounded, while allowing for various heterogeneities in the process across time. Assumption 4.2 holds for all kernel functions commonly used in practical applications. Assumption 4.3 holds as long as $\delta \to 0$ fast enough relative to $T \to \infty$. In particular, if we write $U = U^c + U^d$ where $U^c$ and $U^d$ denote respectively continuous and jump parts of $U$, then the asymptotic modulus of continuity of $U$ is determined by those of $U^c$ and $U^d$ respectively. For any process $U$ with a diffusion type continuous part with a bounded volatility function, and a jump part which has jump size with finite second moment and finite number of jumps that happen in a bounded time interval, we have $\Delta_{\delta,T} = \delta^{1/2}$. Therefore, if we set $B_T = cT^{1/(2r+1)}$ with some constant $c > 0$ and relevant characteristic exponent $r = 1$ or $2$, then the condition requires $\delta^{1/2} T^{1/(2r+1)} \to 0$ as $\delta \to 0$ and $T \to \infty$, which holds if and only if $\delta = o\left(T^{-2/(2r+1)}\right)$. This does not seem to be overly restrictive.\(^{12}\)

**Theorem 4.1.** Let Assumptions 2.1, 2.2, 2.3, 4.1, 4.2 and 4.3 hold. Then we have

$$\Pi_{n,\delta} \to_p \Pi$$

as $\delta \to 0$ and $T \to \infty$.

Theorem 4.1 establishes consistency of the discrete sample LRV estimator $\Pi_{n,\delta}$ of continuous time process. An immediate consequence of our asymptotics here is that the discrete time LRV estimator $\Omega_n$ in (16) does not consistently estimate any well defined parameter, and diverges to infinity in probability as $\delta \to 0$. In fact, this is what we observe from a variety of economic and financial time series at high frequency including the illustrative

\(^{12}\)It is not meaningful to see whether or not this condition holds by directly comparing $\delta$ and $T$, since their comparison is dependent on the time unit. Suppose, for instance, we have daily observations over one year. If the annual time unit is used, we have $\delta \approx 1/250 \ll 1 = T^{-2/(2r+1)}$ and the condition appears to hold. If the daily time unit is used, however, we have $\delta = 1 \gg (1/250)^{2/(2r+1)} \approx T^{-2/(2r+1)}$ and the condition clearly fails to hold.
examples we introduce later in the paper.

There is one important remark to be made here. For consistency of $\Pi_{n,\delta}$, it is absolutely necessary to use a discrete time bandwidth $b_n$ that is given by a continuous time bandwidth $B_T$ as in (17) such that $B_T \to \infty$ and $B_T/T \to 0$. This is usually not the case for a discrete time bandwidth $b_n$ set entirely within a discrete time framework. If, for instance, we set $b_n = cn^p$ with some $c > 0$ and $0 < p < 1$, then we have $B_T = b_n \delta = c\delta^{1-p}T^p \to 0$ whenever $\delta \to 0$ fast enough relative to $T \to \infty$, and therefore, the essential condition $B_T \to \infty$ required for consistency of $\Pi_{n,\delta}$ fails to hold. If $B_T \to 0$ as in this case, we may well expect that

$$\frac{1}{B_T} \Pi_{n,\delta} \approx \frac{1}{B_T} \int_{|s| \leq T} K \left( \frac{s}{B_T} \right) \Gamma_T(s) ds \to_p \Gamma(0) \int_{-\infty}^{\infty} K(x) dx$$

as $\delta \to 0$ and $T \to \infty$, where $\Gamma$ is the autocovariance function of $U$. This implies, in particular, that $\Pi_{n,\delta} \to_p 0$ as $\delta \to 0$ and $T \to \infty$.

Chang and Park (2014) more systematically study asymptotic behaviors of the continuous time bandwidth $B_T$ defined from a discrete time bandwidth $b_n$ as $B_T = b_n \delta$. In their terminology, a discrete time bandwidth $b_n$ is said to be high-frequency compatible if its corresponding continuous time bandwidth $B_T = b_n \delta$ satisfies the critical condition $B_T \to \infty$. As discussed above, it is essential to use a high-frequency compatible discrete time bandwidth $b_n$ for $\Omega_n$ for consistency of $\Pi_{n,\delta} = \delta \Omega_n$. As a part of their investigation on the asymptotics of regressions with observations collected at high frequency over long span, they analyze data-dependent discrete time bandwidth choices such as those in Andrews (1991) and Newey and West (1994). In particular, they show that the semiparametric approach by Andrews (1991) yields a bandwidth $b_n$ whose corresponding continuous time bandwidth defined as $B_T = b_n \delta$ is not only high-frequency compatible but also optimal asymptotically as $\delta \to 0$ and $T \to \infty$. This implies that the estimator $\Pi_{n,\delta} = \delta \Omega_n$ with $\Omega_n$ based on Andrews’ discrete time bandwidth is consistent and asymptotically optimal. In sharp contrast, they find that the discrete time bandwidth $b_n$ obtained from the nonparametric approach by Newey and West (1994) is not high-frequency compatible, and the corresponding continuous time bandwidth $B_T = b_n \delta \to_p 0$. Accordingly, the estimator $\Pi_{n,\delta} = \delta \Omega_n$ using Newey-West’s discrete time bandwidth becomes inconsistent and converges to zero in probability as $\delta \to 0$ and $T \to \infty$.

Finally, we discuss how to obtain a feasible optimal bandwidth using discrete samples. As assumed in Section 3, we let $U$ be stationary. For a nonparametric approach, we may estimate $\Lambda_T(r)$ in (10) using its discrete time version $\Lambda_{n,\delta}$, which is given by

$$\Lambda_{n,\delta}(r) = \delta^{1+r} \sum_{|j| \leq a_n} |j|^r \Gamma_n(j),$$

16
where \( a_n = A_T/\delta \) with \( A_T = cT^p \) for some \( c > 0 \) and \( 0 < p < 1 \) and \( \Gamma_n \) is the sample autocovariance function of \((u_i)\) as defined before. It is easy to see that we have

\[
\Lambda_{n,\delta}(r) = \sum_{|j\delta| \leq a_n \delta} \delta |j\delta| \Gamma_n(j\delta) \approx \int_{|s| \leq A_T} |s|^\tau \Gamma_T(s) = \Lambda_T(r)
\]

as \( \delta \to 0 \) and \( T \to \infty \), under suitable conditions. In the following assumption, we let \( \Delta_{\delta,T} \) be defined as in (19), \( A_T = cT^p \) with some \( c > 0 \) and \( 0 < p < 1 \), and \( r \) be the characteristic exponent of kernel function.

**Assumption 4.4.** \( \Delta_{\delta,T}T^{p(1+r)} \to 0 \) as \( \delta \to 0 \) and \( T \to \infty \).

**Proposition 4.2.** Let Assumptions 2.1, 2.2, 2.3 and 4.4 hold. Then we have

\[
\Lambda_{n,\delta}(r) \to_p \Lambda(r)
\]

as \( \delta \to 0 \) and \( T \to \infty \).

For a semiparametric approach, we let \( U \) be one-dimensional and generated as the general jump diffusion in (12). In this case, we only need to have consistent estimates for its mean reversion parameter \( \kappa \) and variance \( \sigma^2 \). The usual sample variance \( \sigma_{n,\delta}^2 = (1/n) \sum_{i=1}^{n} u_i^2 \) of \((u_i)\) can be used to estimate \( \sigma^2 \). On the other hand, mean reversion parameter \( \kappa \) can be consistently estimated by the usual least squares estimator

\[
\kappa_{n,\delta} = -\left( \frac{1}{\sum_{i=2}^{n} \delta u_{i-1}^2} \right)^{-1} \sum_{i=2}^{n} u_{i-1}(u_i - u_{i-1}),
\]

which corresponds to the continuous time least squares estimator \( \kappa_T \) of \( \kappa \) given by \( \kappa_T = -\left( \int_{0}^{T} U_t^2 dt \right)^{-1} \int_{0}^{T} U_t dU_t \).\(^ {13}\) Below, we introduce some additional technical assumptions to establish consistency of \( \kappa_{n,\delta} \) and \( \sigma_{n,\delta}^2 \) as \( \delta \to 0 \) and \( T \to \infty \).

**Assumption 4.5.** Let \( U \) be generated as in (12), and \((p_T)\) and \((q_T)\) be nonrandom sequences such that

\[
\sup_{0 \leq t \leq T} |U_t| = O_p(p_T), \quad \text{and} \quad \sup_{0 \leq t \leq T} \tau^2(U_t) = O_p(q_T),
\]

and assume

(a) \( \mathbb{E} \left[ U_T^2 \tau^2(U_t) \right] < \infty \), \( \mathbb{E} \left[ U_T^2 \nu^2(U_t) \right]^2 < \infty \),

\(^{13}\)It is also possible to estimate \( \kappa \) by \( \kappa_{n,\delta} = -(1/\delta) \log \alpha_{n,\delta} \) or \( \kappa_{n,\delta} = (1 - \alpha_{n,\delta})/\delta \), where \( \alpha_{n,\delta} = (\sum_{i=1}^{n} u_{i-1}^2)^{-1} \sum_{i=1}^{n} u_{i-1}u_i \). They are all asymptotically equivalent.
(b) $\sum_{0 \leq t \leq T} \mathbb{E} (\Delta U_t)^4 = O(T)$ as $T \to \infty$, and
(c) $\Delta \delta, T P T \to 0$ and $\Delta \delta, T \sqrt{q_T / T} \to 0$, as $\delta \to 0$ and $T \to \infty$.

The conditions introduced in Assumption 4.5 are mild. Part (a) holds, for instance, if $U$ has bounded diffusive and jump volatility functions and has finite variance. The condition in Part (b) is satisfied if the number of jump increases at $T$-rate and jump size has finite fourth moment. Part (c) specifies the rate of $\delta \to 0$ relative to $T \to \infty$. For any jump diffusion $U$ with a bounded volatility function $\tau$, we have $\Delta \delta, T = \delta^{1/2}$ and $q_T = O(1)$. Moreover, if the maximum process of $U$ grows at a rate slower than $O_p(T)$, i.e., $p_T \ll T$, then Part (c) holds as long as $\delta = o(T^{-2})$. This is not restrictive at all.

**Proposition 4.3.** Let Assumptions 4.1 and 4.5 hold. Then we have

$$\kappa_{n, \delta} \to_p \kappa \quad \text{and} \quad \sigma_{n, \delta}^2 \to_p \sigma^2$$

as $\delta \to 0$ and $T \to \infty$.

There are other estimators for mean reversion parameter $\kappa$ that are known to generally perform better. In particular, if $U$ is continuous and has no jump part, the martingale regression method introduced by Park (2013) provides an alternative way to estimate $\kappa$. Moreover, if diffusion and jump terms of $U$ are also known up to any parametric forms, we may use the maximum likelihood estimator of $\kappa$. For the asymptotic theory of maximum likelihood estimator of diffusion and jump diffusion models, the reader is referred respectively to Jeong and Park (2011) and Jeong and Park (2015).

### 5. Illustration and Simulation

To show the practical relevance and usefulness of our continuous time framework and asymptotic theory, we analyze two leading financial time series and perform simulations based on their fitted models. Considered are 1-month forward premium of US/UK exchange rates from January 1, 1979 to June 30, 2015, and 3-month US Treasury bill rates from January 1, 1954 to June 30, 2008, both at daily frequency.\(^{14}\) We assume that they are generated individually by a continuous time process $V$, and that $U$ defined from $V$ as $U = V - \mu$ with

\(^{14}\)The 1-month forward premium is computed as log difference between 1-month forward and spot US/UK exchange rates which are downloaded from the interactive database of Bank of England. The secondary market rate of 3-month US Treasury bill is downloaded from Federal Reserve Economic Data (FRED), Federal Reserve Bank of St. Louis. The most recent T-bill rates after June 30, 2008 are extremely close to the zero lower bound, and are excluded from estimation.
mean $\mu$ of $V$ is given by

\[ \text{OU : } dU_t = -\kappa U_t \, dt + \tau \, dW_t, \]  

(20)

\[ \text{SR : } dU_t = -\kappa U_t \, dt + \tau \sqrt{\mu + U_t} \, dW_t, \]  

(21)

respectively. The process $V = \mu + U$ for $U$ in (20) and (21) is commonly used both in theoretical modeling and empirical application, and referred respectively to as Ornstein-Uhlenbeck (OU) process and Feller’s Square Root (SR) process.\(^\text{15}\) Clearly, both OU and SR processes have linear drifts, and their behaviors rely critically on drift parameter $\kappa$. If $\kappa$ gets larger, their mean reversion gets stronger, and the effect of a shock to them tends to be more transitory. On the contrary, if $\kappa$ is small, the effect of a shock becomes more persistent, making the speed of their mean reversion slower.

We use OU and SR processes $V$ respectively to fit forward premium to sterling and T-bill rates. They yield the parameter estimates $(\mu_0, \kappa_0, \tau_0^2) = (-0.001, 4.826, 0.006^2)$ and $(\mu_0, \kappa_0, \tau_0^2) = (0.052, 0.110, 0.057^2)$ for OU and SR processes.\(^\text{16}\) In our subsequent discussions, we refer to the corresponding $U$ processes as OU-B and SR-B with B signifying the baseline model. To study the effect of mean reversion parameter $\kappa$, we also consider parameter sets given by $(5\kappa_0, 5\tau_0^2)$ and $(\kappa_0/5, \tau_0^2/5)$. These sets of parameter values make $U$ more transitory and persistent respectively without changing its time invariant distribution, and the corresponding process $U$ is referred to similarly as OU-T and OU-P and as SR-T and SR-P, where T and P signify transitory and persistent models, respectively. For both processes, model T and model P have LRVs that are respectively $1/5$ and $5$ times of the LRV of model B.

For our simulations, we use exact transitions of OU and SR processes with initial values drawn from their time invariant distributions. The simulation samples are generated as 2520, 7560, and 12600 daily observations, corresponding to $T = 10, 30$ and $50$ years of sample span. Furthermore, we consider sampling intervals ranging from $1/252$ to $1/2$, which correspond respectively to daily and semi-annual frequencies. The reported means, standard errors and mean squared errors of LRV estimators are computed based on 1000 iterations.

\(^{15}\)The OU process $V = \mu + U$ with $U$ in (20) has time invariant distribution $N(\mu, \tau^2/2\kappa)$ and LRV is given by $\tau^2/\kappa^2$, if $\kappa > 0$ and $\tau > 0$. The SR process $V = \mu + U$ with $U$ in (21) has time invariant distribution given by Gamma distribution with shape parameter $2\kappa\mu/\tau^2$ and rate parameter $2\kappa/\tau^2$, and LRV is given by $\mu\tau^2/\kappa^2$, if $\kappa > 0$, $\tau > 0$ and $2\kappa\mu/\tau^2 \geq 1$.

\(^{16}\)Note that the fitted OU model for forward premium to sterling has a much stronger mean reversion compared to the fitted SR model for T-bill rates.
continuous time LRV estimators, where Bartlett and Parzen kernels are used. For the discrete time LRV estimator, we consider three bandwidth choices which are subsequently referred to as RT, NP and SP. RT stands for rule of thumb and sets $b_n = cn^p$ with $c = 3.7947$ and $p = 1/4$. NP and SP are the data dependent optimal bandwidth choices proposed by Newey and West (1994) and Andrews (1991), which require respectively nonparametric and semiparametric estimations of unknown parameters. For the nonparametric estimation involved in NP, we need another bandwidth parameter, say $a_n$, for which we set $a_n = 4(n/100)^p$ with $p = 2/9$ for Bartlett kernel, and $p = 4/25$ for Parzen kernel as suggested by Newey and West (1994). On the other hand, SP relies on a parametric AR(1) model to estimate an unknown parameter. For the continuous time LRV estimator, we consider three bandwidth choices CRT, CNP and CSP. In particular, we choose $B_T = cT^p$ with $c = 0.5886$ and $p = 1/4$ in CRT, and CNP and CSP are obtained as their feasible versions based on discrete samples. For CNP, we choose first step bandwidth $A_T = 0.5886T^{1/4}$ under CRT.

5.1. Frequency Dependence of Discrete and Continuous Time Estimators

It is well expected from our asymptotic theory that the discrete time LRV estimator $\Omega_n$ diverges as sampling frequency increases. Such a pattern is clearly displayed in Figure 1, where we plot the LRV estimates of forward premium on US/UK exchange rates and US T-bill rates obtained from their observations collected at various sampling intervals. Their LRV estimates vary substantially with sampling intervals. In fact, they increase as sampling frequency increases and, in particular, diverge very rapidly as sampling interval approaches to zero. This is precisely what our asymptotic theory predicts. Figure 2 shows that means of the discrete time LRV estimates from our simulations also have exactly the same patterns. They tend to slowly and steadily increase as sampling interval gets smaller, but explode as sampling interval becomes very small. This tendency generally becomes more noticeable in case the underlying process is more persistent, in both our empirical analysis and simulation study. Note that T-bill rates and our baseline SR model are significantly more persistent than forward premium to sterling and our baseline OU model.

---

17 We only report the results obtained using Parzen kernel here, since there are no meaningful differences in the results obtained from using different kernel functions.

18 Schwert (1989) and Kwiatkowski et al. (1992), in their simulation studies of unit root and stationarity tests, use the rule of thumb bandwidth parameter $b_n = 12(n/100)^{1/4}$, which reduces to our choice of bandwidth with $c = 12/100 = 3.7947$ and $p = 1/4$.

19 As in RT, there is no solid rule to set the constant $c$ in CRT, so we choose it to be comparable to that in RT. In specific, we set $p = 1/4$ as in RT, and $c$ is chosen such that $B_T = cT^p = \delta b_n$ where $b_n = 3.7947n^{1/4}$ under RT scheme, for $\delta = 1/12$ corresponding to the monthly observations used in discrete time studies, and $n = 1000$ which is the largest sample size used in many discrete time LRV estimator simulations (e.g., Newey and West (1994)). Therefore, we use $c = \frac{1/12 \times 3.7947 \times 1000^{1/4}}{(1/4 \times 1000)^{1/4}} = 0.5886$. 

20
Though it converges eventually as sampling frequency increases, the continuous time LRV estimator $\Pi_{n,\delta}$ is also generally quite sensitive to the change in sampling frequency. Figure 3 exhibits how the continuous time LRV estimates for forward premium to sterling and T-bill rates change over different sampling frequencies. To show and analyze the sampling frequency dependence of continuous time LRV estimator more systematically, we present the simulated biases and standard errors obtained from our baseline OU and SR models.\(^20\)

Figure 4 provides the simulated means of biases in continuous time LRV estimators based on discrete samples. Note that the bias consists of two parts: discrete sample bias and continuous sample bias, given respectively by $\mathbb{E}\Pi_{n,\delta} - \mathbb{E}\Pi_T$ and $\mathbb{E}\Pi_T - \Pi$. It follows immediately from Lemma 2.1 and its subsequent discussions that continuous sample bias becomes negative asymptotically for all our simulated models. For our baseline OU model, the biases of continuous time LRV estimators change over different sampling frequencies. Moreover, they are not minimized not at the highest sampling frequency but at some lower sampling frequencies.\(^21\) We believe that this is due to their positive discrete sample biases. In contrast, for the baseline SR model, the magnitude of bias is not affected at all by the sampling frequency. In general, bias depends more on sampling frequency if underlying continuous time model is more transitory. As expected, the overall magnitude of bias decreases especially at high sampling frequencies as sample span increases.

Our simulations show that the standard error of continuous time LRV estimator is also generally dependent upon sampling frequency. See Figure 5. The dependency of its standard error on sampling frequency, however, is not as evident as that of its bias. It is clearly seen only for more transitory baseline OU model and for samples of small spans. As underlying process becomes more persistent or sample span increases, frequency dependence in the standard error of continuous time LRV estimator disappears and becomes unimportant. Moreover, simulated standard errors appear to decrease as sample span increases only for more transitory baseline OU model. For more persistent baseline SR model, such a tendency does not exist and simulated standard errors even increase with sample span in some cases.

There is an additional important issue on the performance of continuous time LRV estimators obtained from discrete samples: the stability of estimators. Both in our empirical analysis and simulation study, continuous time LRV estimators are extremely unstable at lower frequencies, yielding values that change drastically even if we vary sampling frequency only marginally. This is seen clearly in Figures 3 and 4. To further investigate this problem,

\(^{20}\)The simulation results for more transitory and persistent models of both OU and SR processes are have shown very similar patterns as that for the baseline model, so we do not include them.

\(^{21}\)They are positive at low frequencies, but become negative at high frequencies.
we compute the local means and local standard errors of forward premium to sterling and T-bill rates. The local means are defined as two-week local averages around current values, and the local standard errors as two-week local average deviations from the local means. They are presented in Figure 6. The local means are significantly more stable than unaveraged estimates presented in Figure 3. However, they still fluctuate nontrivially in most cases. The local standard errors can indeed be substantial at low sampling frequencies. Our simulation study investigating the instability of continuous time LRV estimators from discrete samples is summarized in Figure 7, where we report the simulated means of two-week local standard errors for our OU and SR simulation models. As expected, the simulated means of local standard errors tend to increase as sampling frequency decreases.

In sum, our empirical analysis and simulation study make it clear that we have a strong incentive to use high frequency observations to estimate continuous time LRVs efficiently. The continuous time LRV estimators using low frequency observations yield nonnegligible biases and serious instabilities. In this context, a question naturally arises regarding how small sampling interval should be to avoid these undesirable biases and instabilities. Doubtless, the answer should be dependent on data and models. For the data and models we use in our empirical analysis and simulation study, daily observations are good enough to obtain continuous time LRV estimators with negligible biases and inconsequential instabilities. In particular, it is clear that we do not need to use ultra-high frequency observations. From our simulations, we find that additional gains from using daily observations in place of weekly observations are already marginal and almost ignorable. Though we do not report the details, using any intra-day observations including ultra-high frequency observations collected at minutes’ intervals does not improve the performance of continuous time LRV estimators in all cases of our simulations.

5.2. Relative Performance of Continuous Time Estimators

In our simulations, we also evaluate the relative performances of continuous time LRV estimator based on different bandwidths. Our comparisons are made in two steps. In the first step, we compare LRV estimators based on bandwidth choices CRT, CNP and CSP. Note that CNP and CSP use asymptotically optimal bandwidths. Of the two, CSP uses more information than CNP, since it further exploits the fact that the underlying continuous time process has a linear drift. In the second step, we focus on CSP implemented with a different estimate for mean reversion parameter. Together with the ordinary least squares estimator (OLS) introduced in the paper, we also consider the maximum likelihood estimator (MLE). Both estimators use the information on linear specification of drift. However, MLE requires a full parametric specification of diffusion, whereas OLS does not require any knowledge on
Table 2 presents the simulated absolute biases, standard errors and mean squared errors of continuous time LRV estimators in variant simulation models using discrete samples collected at daily, monthly and quarterly frequencies over 10, 30 and 50 years of sample span. For all three bandwidth choices, the performance of continuous time LRV estimators generally improves, often significantly, as sample span increases. However, this is not always the case. Indeed, for more persistent SR-B and SR-P models, CSP yields mean squared errors increasing with sample span. On the other hand, using higher frequency observations improves the performance of continuous time LRV estimators only when the underlying process exhibits excessively strong mean reversion as in transitory OU-T model. In all other cases, the continuous time LRV estimators obtained from quarterly observations have close enough, or even comparable, performances as those obtained from monthly or daily observations.

The relative performance of continuous time LRV estimators with bandwidth choices CRT, CNP and CSP can be summarized as follows. Of the three bandwidth choices, CRT produces largest bias with smallest standard errors in all simulation models except for the most transitory OU-T. Also, CSP yields smaller biases and larger standard errors than CNP in all simulation models other than OU-T and OU-B, which represent more transitory underlying processes compared to other simulation models. In terms of mean squared errors, CNP performs best for OU-P, SR-T and SR-B models, CSP for OU-T and SR-P models, and CRT for OU-B model. Therefore, none of three bandwidth choices generally outperforms the others. The comparisons of three bandwidth choices are the same across all sampling frequencies and sample spans considered in our simulations. Finally, the performance of continuous time LRV estimators based on CSP improves when we use MLE to estimate mean reversion parameter in place of OLS in almost all but a few cases under SR-P model, which is excessively persistent. This is shown in Table 3.

6. Conclusion

In this paper, we present the methodology and asymptotic theory for LRV estimation of continuous time process. We analyze the asymptotic bias and variance of LRV estimator and find the optimal bandwidth that minimizes mean squared error of the estimator. Some data dependent bandwidth selection procedures are also developed. Our developments are entirely analogous to those in Andrews (1991), which provides the methodology and asymptotic theory for LRV estimation of discrete time series. Though our methodology and asymptotic theory are motivated and established for continuous time process, the actual
Table 2. Absolute Biases, Standard Deviations and MSEs of Continuous Time LRV Estimators Using CRT, CNP and CSP

<table>
<thead>
<tr>
<th>Bandwidth</th>
<th>CRT</th>
<th>CNP</th>
<th>CSP</th>
<th>CRT</th>
<th>CNP</th>
<th>CSP</th>
<th>CRT</th>
<th>CNP</th>
<th>CSP</th>
<th>CRT</th>
<th>CNP</th>
<th>CSP</th>
<th>CRT</th>
<th>CNP</th>
<th>CSP</th>
<th>CRT</th>
<th>CNP</th>
<th>CSP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily</td>
<td>0.003</td>
<td>0.007</td>
<td>0.014</td>
<td>0.110</td>
<td>0.071</td>
<td>0.074</td>
<td>0.091</td>
<td>0.595</td>
<td>0.292</td>
<td>0.077</td>
<td>0.058</td>
<td>0.025</td>
<td>0.242</td>
<td>0.226</td>
<td>0.160</td>
<td>0.658</td>
<td>0.649</td>
<td>0.596</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.630</td>
<td>2.422</td>
<td>1.035</td>
<td>0.708</td>
<td>1.125</td>
<td>0.951</td>
<td>0.222</td>
<td>0.490</td>
<td>1.076</td>
<td>0.186</td>
<td>0.430</td>
<td>1.218</td>
<td>0.035</td>
<td>0.087</td>
<td>0.364</td>
<td>0.005</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.005</td>
<td>0.010</td>
<td>0.002</td>
<td>0.032</td>
<td>0.057</td>
<td>0.043</td>
<td>0.417</td>
<td>0.295</td>
<td>0.641</td>
<td>0.006</td>
<td>0.005</td>
<td>0.014</td>
<td>0.048</td>
<td>0.043</td>
<td>0.045</td>
<td>0.066</td>
<td>0.064</td>
</tr>
<tr>
<td>Monthly</td>
<td>0.005</td>
<td>0.025</td>
<td>0.056</td>
<td>0.103</td>
<td>0.067</td>
<td>0.055</td>
<td>0.899</td>
<td>0.586</td>
<td>0.287</td>
<td>0.077</td>
<td>0.058</td>
<td>0.024</td>
<td>0.242</td>
<td>0.226</td>
<td>0.160</td>
<td>0.658</td>
<td>0.649</td>
<td>0.596</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>2.057</td>
<td>3.244</td>
<td>1.910</td>
<td>0.713</td>
<td>1.156</td>
<td>1.043</td>
<td>0.223</td>
<td>0.501</td>
<td>1.074</td>
<td>0.187</td>
<td>0.439</td>
<td>1.230</td>
<td>0.035</td>
<td>0.089</td>
<td>0.365</td>
<td>0.005</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.001</td>
<td>0.021</td>
<td>0.009</td>
<td>0.031</td>
<td>0.059</td>
<td>0.048</td>
<td>0.416</td>
<td>0.296</td>
<td>0.636</td>
<td>0.006</td>
<td>0.005</td>
<td>0.014</td>
<td>0.048</td>
<td>0.043</td>
<td>0.045</td>
<td>0.066</td>
<td>0.064</td>
</tr>
<tr>
<td>Quarterly</td>
<td>0.004</td>
<td>0.005</td>
<td>0.016</td>
<td>0.125</td>
<td>0.066</td>
<td>0.082</td>
<td>1.265</td>
<td>0.652</td>
<td>0.321</td>
<td>0.114</td>
<td>0.073</td>
<td>0.031</td>
<td>0.370</td>
<td>0.329</td>
<td>0.139</td>
<td>1.019</td>
<td>0.995</td>
<td>0.765</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.600</td>
<td>2.447</td>
<td>1.010</td>
<td>0.740</td>
<td>1.227</td>
<td>0.937</td>
<td>0.256</td>
<td>0.629</td>
<td>1.067</td>
<td>0.200</td>
<td>0.520</td>
<td>1.193</td>
<td>0.053</td>
<td>0.157</td>
<td>0.920</td>
<td>0.008</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.004</td>
<td>0.010</td>
<td>0.002</td>
<td>0.038</td>
<td>0.066</td>
<td>0.043</td>
<td>0.805</td>
<td>0.410</td>
<td>0.639</td>
<td>0.013</td>
<td>0.008</td>
<td>0.014</td>
<td>0.112</td>
<td>0.093</td>
<td>0.171</td>
<td>0.157</td>
<td>0.150</td>
</tr>
<tr>
<td>Daily</td>
<td>0.004</td>
<td>0.005</td>
<td>0.017</td>
<td>0.124</td>
<td>0.057</td>
<td>0.077</td>
<td>1.474</td>
<td>0.671</td>
<td>0.353</td>
<td>0.135</td>
<td>0.078</td>
<td>0.030</td>
<td>0.451</td>
<td>0.390</td>
<td>0.141</td>
<td>1.249</td>
<td>1.211</td>
<td>0.792</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.651</td>
<td>2.539</td>
<td>1.041</td>
<td>0.798</td>
<td>1.332</td>
<td>1.004</td>
<td>0.267</td>
<td>0.684</td>
<td>1.041</td>
<td>0.233</td>
<td>0.629</td>
<td>1.300</td>
<td>0.060</td>
<td>0.194</td>
<td>1.196</td>
<td>0.010</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.005</td>
<td>0.011</td>
<td>0.002</td>
<td>0.041</td>
<td>0.076</td>
<td>0.047</td>
<td>1.084</td>
<td>0.459</td>
<td>0.621</td>
<td>0.019</td>
<td>0.010</td>
<td>0.016</td>
<td>0.167</td>
<td>0.131</td>
<td>0.278</td>
<td>0.236</td>
<td>0.223</td>
</tr>
<tr>
<td>Quarterly</td>
<td>0.011</td>
<td>0.111</td>
<td>0.108</td>
<td>0.115</td>
<td>0.043</td>
<td>0.050</td>
<td>1.483</td>
<td>0.667</td>
<td>0.343</td>
<td>0.135</td>
<td>0.078</td>
<td>0.030</td>
<td>0.452</td>
<td>0.389</td>
<td>0.141</td>
<td>1.249</td>
<td>1.211</td>
<td>0.792</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>2.074</td>
<td>3.338</td>
<td>1.795</td>
<td>0.800</td>
<td>1.339</td>
<td>1.083</td>
<td>0.263</td>
<td>0.686</td>
<td>1.054</td>
<td>0.229</td>
<td>0.632</td>
<td>1.303</td>
<td>0.059</td>
<td>0.196</td>
<td>1.192</td>
<td>0.010</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.020</td>
<td>0.031</td>
<td>0.001</td>
<td>0.039</td>
<td>0.076</td>
<td>0.051</td>
<td>1.097</td>
<td>0.458</td>
<td>0.632</td>
<td>0.019</td>
<td>0.010</td>
<td>0.016</td>
<td>0.167</td>
<td>0.131</td>
<td>0.277</td>
<td>0.236</td>
<td>0.223</td>
</tr>
<tr>
<td>Daily</td>
<td>0.070</td>
<td>0.716</td>
<td>0.731</td>
<td>0.024</td>
<td>0.052</td>
<td>0.071</td>
<td>1.508</td>
<td>0.654</td>
<td>0.305</td>
<td>0.137</td>
<td>0.077</td>
<td>0.028</td>
<td>0.453</td>
<td>0.389</td>
<td>0.141</td>
<td>1.250</td>
<td>1.211</td>
<td>0.792</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>4.686</td>
<td>7.640</td>
<td>6.123</td>
<td>0.876</td>
<td>1.503</td>
<td>1.402</td>
<td>0.251</td>
<td>0.683</td>
<td>1.090</td>
<td>0.217</td>
<td>0.625</td>
<td>1.297</td>
<td>0.056</td>
<td>0.197</td>
<td>1.185</td>
<td>0.009</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.569</td>
<td>0.611</td>
<td>0.598</td>
<td>0.032</td>
<td>0.096</td>
<td>0.086</td>
<td>1.130</td>
<td>0.448</td>
<td>0.660</td>
<td>0.019</td>
<td>0.009</td>
<td>0.016</td>
<td>0.168</td>
<td>0.131</td>
<td>0.273</td>
<td>0.236</td>
<td>0.223</td>
</tr>
</tbody>
</table>

Notes: Presented are the ratios of simulated absolute biases, standard errors and MSEs of LRV estimators to their asymptotic counterparts of continuous time sample estimators. The latter are obtained according to the small-$b$ asymptotic results in Lemmas 2.1 and 2.2 with stationarity assumption. Results are based on 1000 iterations. Parzen kernel is used in all simulations. We use the least square estimate for mean reversion parameter $\kappa$ to obtain a feasible CSP bandwidth choice.
Table 3. Absolute Biases, Standard Deviations and MSEs of Continuous Time LRV Estimators Using CSP with MLE of $\kappa$

<table>
<thead>
<tr>
<th></th>
<th>OU-T</th>
<th>OU-B</th>
<th>OU-P</th>
<th>SR-T</th>
<th>SR-B</th>
<th>SR-P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily</td>
<td>Bias</td>
<td>0.015</td>
<td>0.075</td>
<td>0.292</td>
<td>0.040</td>
<td>0.190</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.018</td>
<td>0.948</td>
<td>1.076</td>
<td>0.868</td>
<td>0.258</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.002</td>
<td>0.042</td>
<td>0.641</td>
<td>0.008</td>
<td>0.042</td>
</tr>
<tr>
<td>T = 10</td>
<td>Bias</td>
<td>0.051</td>
<td>0.066</td>
<td>0.293</td>
<td>0.040</td>
<td>0.190</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.454</td>
<td>0.980</td>
<td>1.064</td>
<td>0.860</td>
<td>0.258</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.006</td>
<td>0.044</td>
<td>0.628</td>
<td>0.008</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>0.433</td>
<td>0.003</td>
<td>0.282</td>
<td>0.014</td>
<td>0.191</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>4.667</td>
<td>1.141</td>
<td>1.075</td>
<td>0.846</td>
<td>0.254</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.216</td>
<td>0.054</td>
<td>0.637</td>
<td>0.008</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>0.017</td>
<td>0.083</td>
<td>0.321</td>
<td>0.038</td>
<td>0.202</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>0.992</td>
<td>0.934</td>
<td>1.067</td>
<td>1.091</td>
<td>0.726</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.002</td>
<td>0.043</td>
<td>0.639</td>
<td>0.012</td>
<td>0.130</td>
</tr>
<tr>
<td>T = 30</td>
<td>Bias</td>
<td>0.080</td>
<td>0.073</td>
<td>0.322</td>
<td>0.038</td>
<td>0.202</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.318</td>
<td>0.956</td>
<td>1.069</td>
<td>1.092</td>
<td>0.731</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.009</td>
<td>0.043</td>
<td>0.642</td>
<td>0.012</td>
<td>0.131</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>0.646</td>
<td>0.010</td>
<td>0.305</td>
<td>0.038</td>
<td>0.202</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>3.508</td>
<td>1.095</td>
<td>1.083</td>
<td>1.101</td>
<td>0.724</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.438</td>
<td>0.049</td>
<td>0.652</td>
<td>0.012</td>
<td>0.129</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>0.018</td>
<td>0.078</td>
<td>0.353</td>
<td>0.035</td>
<td>0.203</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.022</td>
<td>1.001</td>
<td>1.041</td>
<td>1.229</td>
<td>0.902</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.002</td>
<td>0.044</td>
<td>0.621</td>
<td>0.015</td>
<td>0.183</td>
</tr>
<tr>
<td>T = 50</td>
<td>Bias</td>
<td>0.099</td>
<td>0.066</td>
<td>0.355</td>
<td>0.035</td>
<td>0.203</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.315</td>
<td>1.012</td>
<td>1.040</td>
<td>1.226</td>
<td>0.902</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.013</td>
<td>0.046</td>
<td>0.621</td>
<td>0.015</td>
<td>0.183</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>0.777</td>
<td>0.036</td>
<td>0.338</td>
<td>0.035</td>
<td>0.204</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>3.457</td>
<td>1.172</td>
<td>1.049</td>
<td>1.212</td>
<td>0.890</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.624</td>
<td>0.058</td>
<td>0.624</td>
<td>0.014</td>
<td>0.179</td>
</tr>
</tbody>
</table>

Notes: Presented are the ratios of simulated absolute biases, standard errors and MSEs of LRV estimators using CSP scheme with MLE of $\kappa$ to their asymptotic counterparts of continuous time sample estimators. The latter are obtained according to the small-b asymptotic results in Lemmas 2.1 and 2.2 with stationarity assumption. Results are based on 1000 iterations. Parzen kernel is used in all simulations.
implementation of our procedure is based on discrete samples. Therefore, we make it possible to estimate the LRV of a continuous time process with a feasible data dependent optimal bandwidth choice, relying entirely on a discrete sample from underlying process. It is shown that our estimator is consistent under very mild regularity conditions as sampling interval decreases to zero and time span increases to infinity.

Our framework is also very useful to systematically analyze the frequency dependence of LRV estimators of discrete time series. In particular, our theory developed in the paper implies that, if a discrete sample is collected from a continuous time process, its discrete time LRV estimator diverges up to infinity as sampling interval decreases down to zero. In fact, the tendency for a discrete time LRV estimator to increase as sampling frequency increases is widely seen for many economic and financial time series. This is in sharp contrast with continuous time LRV estimator, which is expected to converge as sampling frequency increases. This indicates that our continuous time framework may be very useful to analyze the frequency dependence of discrete time LRV estimators of many economic and financial time series. Finally, as our simulation shows, for any continuous time process with realistic mean reversion, continuous time LRV can be estimated reasonably well using a sample at low frequency such as quarterly observations. In particular, its estimation does not require a sample at high frequency such as intra-day observations.

References


Chang, Y., Park, J. Y., 2014. Understanding regressions with observations collected at high frequency over long span, working Paper, Indiana University.


Park, J. Y., 2013. Martingale regressions for conditional mean models in continuous time, working Paper, Indiana University.


Appendices

A. Useful Lemmas and Their Proofs

**Lemma A.1.** Let Assumptions 2.1 and 4.2 hold. For \( B_T \to \infty \) and \( B_T/T \to 0 \), we have
(a) \( \sum_{|j| \leq n-1} \int_{j\delta}^{(j+1)\delta} K(s/B_T) ds = O(B_T), \) and
(b) \( \sum_{|j| \leq n-1} \int_{j\delta}^{(j+1)\delta} |K(s/B_T) - K(j\delta/B_T)| ds = O(\delta), \)
as \( \delta \to 0 \) and \( T \to \infty \).

**Proof.** Part (a) follows immediately from
\[
B_T^{-1} \sum_{|j| \leq n-1} \int_{j\delta}^{(j+1)\delta} K(s/B_T) ds = \int_{-T/B_T+\delta/B_T}^{T/B_T} |K(x)| dx \to \int_{-\infty}^{\infty} |K(x)| dx < \infty
\]
as \( \delta \to 0 \) and \( T \to \infty \). To prove Part (b), we assume without loss of generality that \( |K'| \) is monotonically decreasing, by taking \( K'(x) = \sup_{y \geq |x|} |K'(y)| \) instead of \( K' \) if necessary. Then, for any \( j = 0, \ldots, n-1 \) and \( s \in [j\delta, (j+1)\delta] \), we have
\[
|K(s/B_T) - K(j\delta/B_T)| \leq |K'(j\delta/B_T)|(s-j\delta)/B_T \leq \delta B_T^{-1} |K'(j\delta/B_T)|,
\]
from which it follows that
\[
\delta^{-1} \sum_{j=0}^{n-1} \int_{j\delta}^{(j+1)\delta} |K(s/B_T) - K(j\delta/B_T)| ds \leq \delta B_T^{-1} \sum_{j=0}^{n-1} \int_{j\delta}^{(j+1)\delta} |K'(j\delta/B_T)| ds
\]
\[
\leq B_T^{-1} \left[ \delta |K'(0)| + \sum_{j=1}^{n-1} \int_{(j-1)\delta}^{j\delta} |K'(s/B_T)| ds \right]
\]
\[
= \delta B_T^{-1} |K'(0)| + \int_{0}^{T/B_T-\delta/B_T} |K'(x)| dx
\]
\[
\to \int_{0}^{\infty} |K'(x)| dx < \infty
\]
as \( \delta \to 0 \) and \( T \to \infty \). Moreover, for \( j = -(n-1), \ldots, -1 \) and \( s \in [j\delta, (j+1)\delta] \), we have
\[
|K(s/B_T) - K(j\delta/B_T)| \leq |K'(s/B_T)|(s-j\delta)/B_T \leq \delta B_T^{-1} |K'(s/B_T)|,
\]
and therefore,
\[
\delta^{-1} \sum_{j=-(n-1)}^{-1} \int_{j\delta}^{(j+1)\delta} |K(s/B_T) - K(j\delta/B_T)| ds
\]
\[
\leq B_T^{-1} \sum_{j=-(n-1)}^{-1} \int_{j\delta}^{(j+1)\delta} |K'(s/B_T)| ds = \int_{-T/B_T+\delta/B_T}^{0} |K'(x)| dx \to \int_{-\infty}^{0} |K'(x)| dx < \infty
\]
as \( \delta \to 0 \) and \( T \to \infty \). The proof is therefore complete. \( \Box \)
LEMMA A.2. Let Assumption 4.1 hold. We have

(a) sup_{s \in [-T,T]} \mathbb{E}|\Gamma_T(s)| = O(1),
(b) max_{|j| \leq n-1} sup_{s \in [j\delta,(j+1)\delta]} \mathbb{E}|\Gamma(s) - \Gamma_T(j\delta)| = O(\Delta_{\delta,T}), and
(c) max_{|j| \leq n-1} \mathbb{E}|\Gamma_T(j\delta) - \Gamma_n(j)| = O(\Delta_{\delta,T}),
as \delta \to 0 and T \to \infty.

Proof. For the proof of Part (a), we let s > 0 and use Cauchy-Schwarz inequality to deduce that

\[ \mathbb{E}|\Gamma_T(s)| = \mathbb{E}\left| \frac{1}{T} \int_s^T U_t U_{t-s} dt \right| \leq \frac{1}{T} \int_s^T \mathbb{E}|U_t U_{t-s}| dt \leq \frac{1}{T} \int_s^T (\mathbb{E}U_t^2)^{1/2} (\mathbb{E}U_{t-s}^2)^{1/2} dt. \]

Similarly, for s \leq 0, we have

\[ \mathbb{E}|\Gamma_T(s)| \leq \frac{1}{T} \int_0^{T-|s|} (\mathbb{E}U_t^2)^{1/2} (\mathbb{E}U_{t-s}^2)^{1/2} dt. \]

Therefore, it follows from Assumption 4.1 that

\[ \sup_{s \in [-T,T]} \mathbb{E}|\Gamma_T(s)| \leq \frac{1}{T} \int_0^T \left( \sup_{0 \leq t \leq T} \mathbb{E}U_t^2 \right) dt = \sup_{0 \leq t \leq T} \mathbb{E}U_t^2 = O(1) \]
as T \to \infty, as was to be shown.

To prove Part (b), we write

\[ \Gamma_T(s) - \Gamma_T(j\delta) = \frac{1}{T} \int_s^T U_t(U_{t-s} - U_{t-j\delta}) dt - \frac{1}{T} \int_j^{j\delta} U_t U_{t-j\delta} dt \]
for j \geq 0 and s \in [j\delta,(j+1)\delta], and

\[ \Gamma_T(s) - \Gamma_T(j\delta) = \frac{1}{T} \int_0^{T-|j\delta|} U_t(U_{t-s} - U_{t-j\delta}) dt - \frac{1}{T} \int_{T-|j\delta|}^{T-|s|} U_t U_{t-s} dt \]
for j < 0 and s \in [j\delta,(j+1)\delta]. It follows from Cauchy Schwarz inequality that

\[ \mathbb{E}|U_t(U_{t-s} - U_{t-j\delta})| \leq (\mathbb{E}U_t^2)^{1/2} (\mathbb{E}U_{t-s}^2)^{1/2} \leq \left( \sup_{0 \leq t \leq T} \mathbb{E}U_t^2 \right)^{1/2} \left[ \sup_{0 \leq s, t \leq T} \sup_{|t-s| \leq \delta} \mathbb{E}(U_t - U_s)^2 \right]^{1/2} \]
for any 0 \leq t \leq T and s \in [j\delta,(j+1)\delta], and

\[ \mathbb{E}|U_t U_{t-s}| \leq (\mathbb{E}U_t^2)^{1/2} (\mathbb{E}U_{t-s}^2)^{1/2} \leq \sup_{0 \leq t \leq T} \mathbb{E}U_t^2 \]
for any $0 \leq t, s \leq T$. Therefore, we may deduce from (22) and (23) that

$$
\max_{|j| \leq n-1} \sup_{s \in [j\delta, (j+1)\delta]} \mathbb{E}[\Gamma_T(s) - \Gamma_T(j\delta)] \leq \left( \sup_{0 \leq t \leq T} \mathbb{E}U_t^2 \right)^{1/2} \left[ \sup_{0 \leq s, t \leq T} \mathbb{E}(U_t - U_s)^2 \right]^{1/2} + \frac{\delta}{T} \sup_{0 \leq t \leq T} \mathbb{E}U_t^2
$$

$$
= O(1)O(\Delta_{\delta,T}) + \delta/TO(1) = O(\Delta_{\delta,T})
$$
as $\delta \to 0$ and $T \to \infty$, as desired.

Lastly, for Part (c), we only consider $j \geq 0$, since the proof for $j < 0$ is entirely analogous. We write

$$
\Gamma_T(j\delta) - \Gamma_n(j) = \frac{1}{T} \sum_{i=j+1}^{n} \int_{(i-1)\delta}^{i\delta} (U_t - U_i) dt = A_{nj} + B_{nj},
$$

where

$$
A_{nj} = \frac{1}{T} \sum_{i=j+1}^{n} \int_{(i-1)\delta}^{i\delta} U_i(U_t - U_i) dt
$$

$$
B_{nj} = \frac{1}{T} \sum_{i=j+1}^{n} \int_{(i-1)\delta}^{i\delta} (U_t - U_i)U_i dt.
$$

Due to Cauchy-Schwarz inequality, we have

$$
\mathbb{E}|U_t(U_t - U_i)| \leq \left( \mathbb{E}U_t^2 \right)^{1/2} \left[ \mathbb{E}(U_t - U_i)^2 \right]^{1/2}
$$

for $i = j+1, \ldots, n$ and $t \in [(i-1)\delta, i\delta]$, from which and (25) it follows that

$$
\max_{0 \leq j \leq n-1} \mathbb{E}|A_{nj}| \leq \max_{0 \leq j \leq n-1} \frac{1}{T} \sum_{i=j+1}^{n} \mathbb{E}|U_i(U_t - U_i)| dt
$$

$$
\leq \left( \sup_{0 \leq t \leq T} \mathbb{E}U_t^2 \right)^{1/2} \left[ \sup_{0 \leq s, t \leq T} \mathbb{E}(U_t - U_s)^2 \right]^{1/2} \frac{1}{T} \int_0^T dt = O(\Delta_{\delta,T}).
$$

Similarly, we may deduce that $\max_{0 \leq j \leq n-1} \mathbb{E}|B_{nj}| = O(\Delta_{\delta,T})$. Therefore, we deduce from (24) that

$$
\max_{0 \leq j \leq n-1} \mathbb{E}|\Gamma_T(j\delta) - \Gamma_n(j)| \leq \max_{0 \leq j \leq n-1} \mathbb{E}|A_{nj}| + \max_{0 \leq j \leq n-1} \mathbb{E}|B_{nj}| = O(\Delta_{\delta,T})
$$
as $\delta \to 0$ and $T \to \infty$. This was to be shown.

\[\square\]

B. Proofs of Theorems

Proof of Lemma 2.1. It is clear that we may assume without loss of generality that $U$ is univariate. For a scalar process $U$, we denote $\varpi$ and $\varpi_T$ as its LRV and continuous time LRV estimator.
respectively, and let $\gamma_T$ be the univariate counterpart of $\Gamma_T$. The bias of $\bar{\omega}_T$ is

$$E\bar{\omega}_T - \omega = \int_{-T}^{T} K(s/B_T)E\gamma_T(s)ds - \lim_{T \to \infty} \int_{-T}^{T} E\gamma_T(s)ds. \quad (27)$$

We show the results under small-$b$ asymptotics and fixed-$b$ asymptotics separately.

(i) Small-$b$ asymptotics

Under small-$b$ asymptotics where $B_r T / T \to 0$ as $T \to \infty$, we multiply $B_T^r$ on both sides of (27) and rearrange terms to get

$$B_T^r (E\bar{\omega}_T - \omega) = -(P_T + Q_T), \quad (28)$$

where

$$P_T = B_T^r \int_{-T}^{T} [1 - K(s/B_T)]E\gamma_T(s)ds \quad \text{and} \quad Q_T = B_T^r \int_{|s| \geq T} E\gamma_T(s)ds.$$ 

For $P_T$, we have

$$|P_T| \leq B_T^r \int_{-T}^{T} [1 - K(s/B_T)] \left( T - \frac{|s|}{T} \sup_{t \geq 0} |E(U_t U_{t-s})| \right) ds = P_{1T} + P_{2T}, \quad (29)$$

where

$$P_{1T} = B_T^r \int_{-T}^{T} [1 - K(s/B_T)] \sup_{t \geq 0} |E(U_t U_{t-s})| ds$$

and

$$P_{2T} = B_T^r \int_{|s| \geq T} [1 - K(s/B_T)] |s| \sup_{t \geq 0} |E(U_t U_{t-s})| ds.$$ 

However, we have

$$P_{1T} = \int_{-T}^{T} \frac{1 - K(s/B_T)}{|s/B_T|^r} |s|^r \sup_{t \geq 0} |E(U_t U_{t-s})| ds \to \pi(r) \int_{-\infty}^{\infty} |s|^r \sup_{t \geq 0} |E(U_t U_{t-s})| ds$$

as $T \to \infty$ by dominated convergence, and

$$P_{2T} \leq \frac{B_T^r}{T} \int_{-T}^{T} |s| \sup_{t \geq 0} |E(U_t U_{t-s})| ds \leq \frac{B_T^r}{T} \int_{-\infty}^{\infty} |s| \sup_{t \geq 0} |E(U_t U_{t-s})| ds \to 0$$

as $T \to \infty$ since $B_T^r / T \to 0$ as $T \to \infty$. Therefore, it follows that

$$\limsup_{T \to \infty} |P_T| \leq \pi(r) \int_{-\infty}^{\infty} |s|^r \sup_{t \geq 0} |E(U_t U_{t-s})| ds. \quad (30)$$

from (29).

For $Q_T$, we have

$$|Q_T| \leq B_T^r \int_{|s| \geq T} \frac{T - |s|}{T} \sup_{t \geq 0} |E(U_t U_{t-s})| ds = Q_{1T} + Q_{2T}, \quad (31)$$
where

\[ Q_{1T} = B_T^r \int_{|s| \geq T} \sup_{t \geq 0} |\mathbb{E}(U_t U_{t-s})| ds \quad \text{and} \quad Q_{2T} = \frac{B_T^r}{T} \int_{|s| \geq T} |s| \sup_{t \geq 0} |\mathbb{E}(U_t U_{t-s})| ds. \]

Note that

\[ Q_{1T} = \int_{|s| \geq T} |B_T/s|^r |s|^r \sup_{t \geq 0} |\mathbb{E}(U_t U_{t-s})| ds \leq \int_{|s| \geq T} |s|^r \sup_{t \geq 0} |\mathbb{E}(U_t U_{t-s})| ds \to 0 \]

as \( T \to \infty \), and \( Q_{2T} \to 0 \) since \( B_T^r/T \to 0 \) as \( T \to \infty \). Therefore, we have

\[ \lim_{T \to \infty} \sup |Q_T| = 0, \quad (32) \]

Consequently, we may deduce from (28), (30) and (32) that

\[ \lim_{T \to \infty} \sup \frac{B_T^r}{T} |\mathbb{E}\omega_T - \omega| \leq \pi(r) \int_{-\infty}^{\infty} |s|^r \sup_{t \geq 0} |\mathbb{E}(U_t U_{t-s})|, \]

as was to be shown.

(ii) Fixed-\( b \) asymptotics

Under fixed-\( b \) asymptotics, we multiply \( B_T \) on both sides of (27) and obtain the similar expression as in (28) with \( r \) replaced by 1. In what follows, we redefine \( P_T \) and \( Q_T \) and their decompositions in (29) and (31) with \( r \) replaced by 1. Then we have

\[ P_{1T} = \int_{-T}^{T} \frac{1 - K(s/B_T)}{|s/B_T|} |s| \sup_{t \geq 0} |\mathbb{E}(U_t U_{t-s})| ds \to \pi(1) \int_{-\infty}^{\infty} |s| \sup_{t \geq 0} |\mathbb{E}(U_t U_{t-s})| ds \]

as \( T \to \infty \), and

\[ P_{2T} \leq \frac{B_T}{T} \int_{-T}^{T} |s| \sup_{t \geq 0} |\mathbb{E}(U_t U_{t-s})| ds \to b \int_{-\infty}^{\infty} |s| \sup_{t \geq 0} |\mathbb{E}(U_t U_{t-s})| ds \]

as \( T \to \infty \). Therefore, it follows that

\[ \lim_{T \to \infty} \sup |P_T| \leq |\pi(1) + b| \int_{-\infty}^{\infty} |s| \sup_{t \geq 0} |\mathbb{E}(U_t U_{t-s})| ds. \quad (33) \]

Moreover, we have

\[ Q_{1T} = \int_{|s| \geq T} |B_T/s|^s |s| \sup_{t \geq 0} |\mathbb{E}(U_t U_{t-s})| ds \leq \int_{|s| \geq T} |s| \sup_{t \geq 0} |\mathbb{E}(U_t U_{t-s})| ds \to 0 \]

as \( T \to \infty \), and

\[ Q_{2T} = \frac{B_T}{T} \int_{|s| \geq T} |s| \sup_{t \geq 0} |\mathbb{E}(U_t U_{t-s})| ds = b \int_{|s| \geq T} |s| \sup_{t \geq 0} |\mathbb{E}(U_t U_{t-s})| ds \to 0 \]
as $T \to \infty$, from which it follows that
\[
\limsup_{T \to \infty} |Q_T| = 0. \tag{34}
\]
Finally, we deduce from (33) and (34) that
\[
\limsup_{T \to \infty} B_T |E \varpi_T - \varpi| \leq [\pi(1) + b] \int_{-\infty}^{\infty} |s| \sup_{t \geq 0} |E(U_t U_{t-s})|,
\]
which completes the proof.

**Proof of Lemma 2.2.** Without loss of generality, we let $U$ be univariate. Using the notations introduced in the proof of Lemma 2.1, the variance of $\varpi_T$ is given as
\[
\var(\varpi_T) = E(\varpi_T - E \varpi_T)^2 = E \left[ \int_{-T}^{T} K(s/B_T) F_T(s) ds \right]^2 \tag{35}
\]
with
\[
F_T(s) = \gamma_T(s) - E\gamma_T(s) = \frac{1}{T} \int_{I_T(s)} [U_t U_{t-s} - E(U_t U_{t-s})] dt \tag{36}
\]
where the integration is taking over
\[
I_T(s) = \begin{cases} 
0, & s \leq 0, \\
[s, T], & s > 0.
\end{cases}
\]
Moreover, we write (35) as
\[
\var(\varpi_T) = \int_{-T}^{T} \int_{-T}^{T} K(s_1/B_T)K(s_2/B_T)E[F_T(s_1)F_T(s_2)] ds_1 ds_2. \tag{37}
\]
Note that we have
\[
F_T(s_1)F_T(s_2) = \frac{1}{T^2} \left( \int_{I_T(s_1)} [U_{t_1} U_{t_1-s_1} - E(U_{t_1} U_{t_1-s_1})] dt_1 \right) \left( \int_{I_T(s_2)} [U_{t_2} U_{t_2-s_2} - E(U_{t_2} U_{t_2-s_2})] dt_2 \right)
= \frac{1}{T^2} \int_{I_T(s_1)} \int_{I_T(s_2)} [U_{t_1} U_{t_1-s_1} - E(U_{t_1} U_{t_1-s_1})][U_{t_2} U_{t_2-s_2} - E(U_{t_2} U_{t_2-s_2})] dt_1 dt_2,
\]
on both sides of which we take expectations and get
\[
E[F_T(s_1)F_T(s_2)] = \frac{1}{T^2} \int_{I_T(s_1)} \int_{I_T(s_2)} [E(U_{t_1} U_{t_1-s_1} U_{t_2} U_{t_2-s_2})]
- E(U_{t_1} U_{t_1-s_1})E(U_{t_2} U_{t_2-s_2})] dt_1 dt_2. \tag{38}
\]
Now we denote the fourth order moment of $U$ by $C(t_1, t_2, s_1, s_2) = E(U_{t_1} U_{t_1-s_1} U_{t_2} U_{t_2-s_2})$, and let $U^\ast$ be the zero mean Gaussian process with the same covariance function as $U$. Then the fourth order moment of $U^\ast$ is given as
\[
C^\ast(t_1, t_2, s_1, s_2) = E(U^\ast_{t_1} U^\ast_{t_1-s_1} U^\ast_{t_2} U^\ast_{t_2-s_2})
= E(U_{t_1} U_{t_1-s_1})E(U_{t_2} U_{t_2-s_2}) + E(U_{t_1} U_{t_2})E(U_{t_1-s_1} U_{t_2-s_2}) + E(U_{t_1} U_{t_2-s_2})E(U_{t_1-s_1} U_{t_2}).
\]
Furthermore, we let $G(t_1, t_2, s_1, s_2)$ be given as

$$G(t_1, t_2, s_1, s_2) = C(t_1, t_2, s_1, s_2) - C^*(t_1, t_2, s_1, s_2),$$

and deduce that

$$E(U_t U_{t+s_1} U_{t+s_2}) = G(t_1, t_2, s_1, s_2) + E(U_t U_{t+s_1})E(U_{t+s_2}) + E(U_t U_{t+s_2})E(U_{t+s_1}) + E(U_t U_{t+s_1+s_2})E(U_{t+s_1+s_2}).$$

(39)

Substituting (39) into (38), and by the change of variables $s = t_2 - t_1$, we have

$$E[F_T(s_1)F_T(s_2)] = \frac{1}{T^2} \int_{I_T(s_2)} \int_{I_T(s_1)} |G(t_1, t_2, s_1, s_2) + E(U_t U_{t+s_1})E(U_{t+s_2}) + E(U_t U_{t+s_2})E(U_{t+s_1}) dt_1 dt_2$$

$$= \frac{1}{T^2} \int_T^{T} \int_{0}^{T} U_T(t, s, s_1, s_2)W(t, s, s_1, s_2) dt ds,$$

where $U_T$ is an indicator function such that $|U_T| \leq 1$, and

$$W(t, s, s_1, s_2) = G(t, t + s, s_1, s_2)$$

$$+ E(U_t U_{t+s})E(U_{t+s_1} U_{t+s_2}) + E(U_t U_{t+s_2})E(U_{t+s_1} U_{t+s}).$$

(41)

Therefore, we may deduce from (40) and (41) that

$$|E(F_T(s_1)F_T(s_2))| \leq \frac{1}{T} \left[ \int_{-T}^{T} \sup_{t \geq 0} |G(t, t + s, s_1, s_2)| ds$$

$$+ \int_{-T}^{T} \sup_{t \geq 0} |E(U_t U_{t+s})| \sup_{t \geq 0} |E(U_t U_{t+s_1+s_2})| ds$$

$$+ \int_{-T}^{T} \sup_{t \geq 0} |E(U_t U_{t+s_2})| \sup_{t \geq 0} |E(U_t U_{t+s_1})| ds \right],$$

by which and (37) have

$$\frac{T}{B_T} \text{var}(\varphi_T) \leq P_T + Q_T + R_T,$$

(42)

where

$$P_T = \frac{1}{B_T} \int_T^{T} \int_{-T}^{T} |K(s_1/B_T)K(s_2/B_T)| \left( \int_{-T}^{T} \sup_{t \geq 0} |G(t, t + s, s_1, s_2)| ds \right) ds_1 ds_2$$

$$Q_T = \frac{1}{B_T} \int_T^{T} \int_{-T}^{T} |K(s_1/B_T)K(s_2/B_T)| \left( \int_{-T}^{T} \sup_{t \geq 0} |E(U_t U_{t+s})| \sup_{t \geq 0} |E(U_t U_{t+s_1+s_2})| ds \right) ds_1 ds_2$$

$$R_T = \frac{1}{B_T} \int_T^{T} \int_{-T}^{T} |K(s_1/B_T)K(s_2/B_T)| \left( \int_{-T}^{T} \sup_{t \geq 0} |E(U_t U_{t+s_2})| \sup_{t \geq 0} |E(U_t U_{t+s_1})| ds \right) ds_1 ds_2.$$

In what follows, we analyze $P_T$, $Q_T$, and $R_T$ respectively under both small-$b$ and fixed-$b$ asymp-
asymptotics. First, since $B_T \to \infty$ as $T \to \infty$ under both asymptotics, we have
\[
P_T \leq B_T^{-1} \int_{-T}^{T} \int_{-T}^{T} \sup_{t \geq 0} |G(t, t + s, s_1, s_2)| dsds_1ds_2 \to 0 \quad (43)
\]
as $T \to \infty$ under Assumption 2.3 (b). Next we consider $Q_T$. Note that by the change of variables $Z_1 = s_1 - s_2$ and $Z_2 = s_2/B_T$, we have
\[
Q_T = B_T^{-1} \int_{-T}^{T} ds_2 \int_{-T}^{T} ds_1 \int_{-T}^{T} ds \left[ |K((Z_1 + s_2)/B_T)K(s_2/B_T)| \sup_{t \geq 0} |\mathbb{E}(U_t U_{t+s})| \sup_{t \geq 0} |\mathbb{E}(U_t U_{t+s+z_1})| \right] \\
= \int_{-T/B_T}^{T/B_T} \int_{-T(1-Z_2/B_T)}^{T(1-Z_2/B_T)} dz \int_{-T}^{T} ds_1 \int_{-T}^{T} ds \left[ |K((Z_1/B_T + z)K(z)| \sup_{t \geq 0} |\mathbb{E}(U_t U_{t+s})| \sup_{t \geq 0} |\mathbb{E}(U_t U_{t+s+z_1})| \right],
\]
from which we may deduce that under small-$b$ asymptotics where $T/B_T \to \infty$,
\[
Q_T \to \left( \int_{-\infty}^{\infty} \sup_{t \geq 0} |\mathbb{E}(U_t U_{t+s})| |ds \right)^2 \int_{-\infty}^{\infty} K^2(x) dx \quad (44)
\]
as $T \to \infty$, whereas under fixed-$b$ asymptotics where $B_T/T = b$, we have
\[
Q_T \to \left( \int_{-\infty}^{\infty} \sup_{t \geq 0} |\mathbb{E}(U_t U_{t+s})| |ds \right)^2 \int_{-1/b}^{1/b} K^2(x) dx \quad (45)
\]
as $T \to \infty$. For $R_T$, we change variables by $Z_2 = s_2 - s$, $Z_1 = s_1 + s_2$, and $Z_2 = s_2/B_T$, and obtain
\[
R_T = \int_{-T/B_T}^{T/B_T} dZ_2 \int_{T(Z_2/B_T/T-1)}^{T(Z_2/B_T/T+1)} dZ_1 \left[ |K((Z_1/B_T - Z_2)K(Z_2)| \sup_{t \geq 0} |\mathbb{E}(U_t U_{t-z})| \sup_{t \geq 0} |\mathbb{E}(U_t U_{t+z_1-z})| \right] \\
\]
from which we deduce
\[
R_T \to \left( \int_{-\infty}^{\infty} \sup_{t \geq 0} |\mathbb{E}(U_t U_{t+s})| |ds \right)^2 \int_{-\infty}^{\infty} K^2(x) dx \quad (46)
\]
as $T \to \infty$ under small-$b$ asymptotics, and
\[
R_T \to \left( \int_{-1/b}^{1/b} \sup_{t \geq 0} |\mathbb{E}(U_t U_{t+s})| |ds \right)^2 \int_{-\infty}^{\infty} K^2(x) dx \quad (47)
\]
as $T \to \infty$ under fixed-$b$ asymptotics where $B_T/T = b$.

Finally, by (42), (43), (44) and (46), we have
\[
\limsup_{T \to \infty} \frac{T}{B_T} \text{var}(\omega_T) \leq 2 \left( \int_{-\infty}^{\infty} \sup_{t \geq 0} |\mathbb{E}(U_t U_{t+s})| |ds \right)^2 \int_{-\infty}^{\infty} K^2(x) dx
\]
under small-$b$ asymptotics. While under fixed-$b$ asymptotics, it follows from (42), (43), (45), (47),
and the fact $T/B_T = 1/b$ that

$$\limsup_{T \to \infty} \text{var}(\omega_T) \leq 2b \left( \int_{-\infty}^{\infty} \sup_{t \geq 0} |\mathbb{E}(U_t U_{t+s})| ds \right)^2 \int_{-1/b}^{1/b} K^2(x) dx.$$ 

This completes the proof. \hfill \Box

**Proof of Theorem 2.3.** By Lemma 2.1 and 2.2, we have for any vector $v \in \mathbb{R}^d$,

$$\limsup_{T \to \infty} \frac{T}{B_T} \mathbb{E}[v'(\Pi_T - \Pi)v]^2 = \limsup_{T \to \infty} \frac{T}{B_T} \mathbb{E}[v'(\Pi_T - \Pi)v]^2 + \limsup_{T \to \infty} \frac{T}{B_T} \text{var}(v'\Pi_T v)$$

$$= \limsup_{T \to \infty} \frac{B_T^2 \mathbb{E}[v'(\Pi_T - \Pi)v]^2}{\lambda + o(1)} + \limsup_{T \to \infty} \frac{T}{B_T} \text{var}(v'\Pi_T v) = O(1),$$

for some $\lambda > 0$ such that $B_T^{2\alpha+1}/T \to \lambda$. Therefore, under small-$b$ asymptotics,

$$\mathbb{E}[v'(\Pi_T - \Pi)v]^2 \to 0$$

as $T \to \infty$, and hence $v'(\Pi_T - \Pi)v = o_p(1)$, for any $v \in \mathbb{R}^d$. This shows consistency of $\Pi_T$ under small-$b$ asymptotics.

Now let $B_T = bT$ with $b > 0$ under fixed-$b$ asymptotics. We may write $\Pi_T$ as

$$\Pi_T = \frac{1}{T} \int_0^T \int_0^T K \left( \frac{t-s}{bT} \right) U_t U_s' dt ds$$

$$= \frac{1}{T} \int_0^T \int_0^T k(t,s) U_t U_s' dt ds = \frac{1}{T} \int_0^T \left( \int_0^t k(t,s) U_t ds \right) U_s' dt,$$

where we define $k(t,s) = K((t-s)/bT)$. Therefore, if we let

$$N_s = \int_0^T k(t,s) U_t dt = \int_0^T k(t,s) d \left( \int_0^t U_r dr \right)$$

$$= \sqrt{T} \int_0^1 k(tT,s) d \left( \frac{1}{\sqrt{T}} \int_0^{tT} U_r dr \right) = \sqrt{T} \int_0^1 k(tT,s) d U_T^s,$$

where $U_T$ is defined in (1), then it follows that

$$\Pi_T = \frac{1}{T} \int_0^T N_s U_s' ds = \frac{1}{T} \int_0^T N_s d \left( \int_0^s U_r' dr \right)$$

$$= \frac{1}{\sqrt{T}} \int_0^1 N_s d \left( \frac{1}{\sqrt{T}} \int_0^{sT} U_r' dr \right) = \frac{1}{\sqrt{T}} \int_0^1 N_s d U^T_s$$

$$= \int_0^1 \int_0^1 k(tT,sT) d U'_T d U^T_s = \int_0^1 \int_0^1 K \left( \frac{t-s}{b} \right) d U'_t d U^T_s.$$

Consequently, by the continuous time version of Donsker’s invariance principle (2) and continuous mapping theorem, we have

$$\Pi_T \to_d \int_0^1 \int_0^1 K \left( \frac{t-s}{b} \right) d U_t^o d U^o_s,$$

37
as $T \to \infty$, as desired to be shown. \hfill \square

**Proof of Lemma 3.1.** For $U$ defined as in (12), we define $\phi : \mathbb{R}^2 \to \mathbb{R}$ as

$$\phi(x,t) = E(U_t|U_0 = x).$$

Note that $U$ is a Markov process with infinitesimal generator $A$ given as

$$Af(x) = -\kappa x \frac{\partial}{\partial x} f(x) + \frac{1}{2} \sigma^2(x) \frac{\partial^2}{\partial x^2} f(x) + \int_{\mathbb{R}} [f(x+z) - f(x)] \nu(dz),$$

where $\nu$ is the Lévy measure of jump process $J$ in (12). Then by Dynkin’s formula (Theorem 7.4.1 in Øksendal (2013)), $\phi(x,t)$ is the unique solution to the Kolmogorov backward equation

$$\frac{\partial}{\partial t} \phi(x,t) = A\phi(x,t)$$

$$= -\kappa x \frac{\partial}{\partial x} \phi(x,t) + \frac{1}{2} \sigma^2(x) \frac{\partial^2}{\partial x^2} \phi(x,t) + \int_{\mathbb{R}} [\phi(x+z,t) - \phi(x,t)] \nu(dz) \quad (48)$$

with initial condition $\phi(x,0) = x$. Given $\int z \nu(dz) = 0$, we can easily check that

$$\phi(x,t) = e^{-\kappa t} x$$

is a solution to (48), so it is the unique solution. Therefore, we may deduce that

$$\Gamma(s) = E(U_{t+s}U_t) = E[U_tE(U_{t+s}|U_t)] = E(e^{-\kappa s}U_t^2) = e^{-\kappa s} \sigma^2$$

for $s \geq 0$, and the proof is complete upon noticing that $\Gamma(s) = \Gamma(-s)$ for $s < 0$. \hfill \square

**Proof of Proposition 3.2.** When each component $U_i$ of $U$ is generated as in (12) with mean reversion parameter $\kappa_i$ and variance $\sigma_i^2$ for $i = 1, \ldots, d$, it follows from Lemma 3.1 that the autocovariance function of $U_i$ is given by $\Gamma_i(s) = e^{-\kappa_i |s|} \sigma_i^2$. Therefore it can be easily deduced that for any positive integer $r$,

$$\int_{-\infty}^{\infty} \Gamma_i(s) ds = \frac{2\sigma_i^2}{\kappa_i} \quad \text{and} \quad \int_{-\infty}^{\infty} |s| \Gamma_i(s) ds = \frac{2r!\sigma_i^2}{\kappa_i^{r+1}}. \quad (49)$$

Then Proposition 3.2 follows immediately by substituting (49) into (11). \hfill \square

**Proof of Theorem 4.1.** Clearly, it suffices to show that

$$\Pi_T - \Pi_{n,\delta} = o_p(1)$$

as $\delta \to 0$ and $T \to \infty$. We write

$$\Pi_T - \Pi_{n,\delta} = P_{n,\delta} + Q_{n,\delta} + R_{n,\delta},$$

where $P_{n,\delta}$ represents the pricing error due to discretization, $Q_{n,\delta}$ is the approximation error of the quadratic approximation, and $R_{n,\delta}$ captures other sources of error. The remaining details are left as an exercise to the reader.
where

$$P_{n,\delta} = \delta \sum_{|j| \leq n-1} K(j\delta/B_T)[\Gamma_T(j\delta) - \Gamma_n(j)]$$

$$Q_{n,\delta} = \sum_{|j| \leq n-1} \int_{j\delta}^{(j+1)\delta} [K(s/B_T)\Gamma_T(s) - K(j\delta/B_T)\Gamma_T(j\delta)]ds$$

$$R_{n,\delta} = \int_{-n\delta}^{(n-1)\delta} K(s/B_T)\Gamma_T(s)ds,$$

each of which will be analyzed in the sequel.

For $P_{n,\delta}$, we have

$$E|P_{n,\delta}| \leq \delta \sum_{|j| \leq n-1} |K(j\delta/B_T)||E[\Gamma_T(j\delta) - \Gamma_n(j)]|
\leq \left( \max_{|j| \leq n-1} |E[\Gamma_T(j\delta) - \Gamma_n(j)]| \right) \delta \sum_{|j| \leq n-1} |K(j\delta/B_T)| = O(\Delta_{\delta,T}B_T),$$

by Lemma A.1 and Lemma A.2 (c). Therefore, $P_{n,\delta} = o_p(1)$ under Assumption 4.3.

To analyze $Q_{n,\delta}$, we write

$$Q_{n,\delta} = F_{n,\delta} + G_{n,\delta}$$

where

$$F_{n,\delta} = \sum_{|j| \leq n-1} \int_{j\delta}^{(j+1)\delta} K(s/B_T)[\Gamma_T(s) - \Gamma_T(j\delta)]ds$$

$$G_{n,\delta} = \sum_{|j| \leq n-1} \int_{j\delta}^{(j+1)\delta} [K(s/B_T) - K(j\delta/B_T)]\Gamma_T(j\delta)ds.$$

For $F_{n,\delta}$, we have

$$E|F_{n,\delta}| \leq \sum_{|j| \leq n-1} \int_{j\delta}^{(j+1)\delta} |K(s/B_T)||E[\Gamma_T(s) - \Gamma_T(j\delta)]|ds
\leq \left( \max_{|j| \leq n-1} \sup_{s \in [j\delta,(j+1)\delta]} E[\Gamma_T(s) - \Gamma_T(j\delta)] \right) \sum_{|j| \leq n-1} \int_{j\delta}^{(j+1)\delta} |K(s/B_T)|ds = O(\Delta_{\delta,T}B_T),$$

(51)

by Lemma A.1 (a) and Lemma A.2 (b). For $G_{n,\delta}$, we notice that

$$E|G_{n,\delta}| \leq \sum_{|j| \leq n-1} \int_{j\delta}^{(j+1)\delta} |K(s/B_T) - K(j\delta/B_T)||E[\Gamma_T(j\delta)]|ds
\leq \left( \max_{|j| \leq n-1} E[\Gamma_T(j\delta)] \right) \sum_{|j| \leq n-1} \int_{j\delta}^{(j+1)\delta} |K(s/B_T) - K(j\delta/B_T)|ds = O(\delta),$$

(52)

39
by Lemma A.1 (b) and Lemma A.2 (a). Therefore, it follows from (50), (51) and (52) that

$$Q_{n,\delta} = O_p(\Delta_{\delta,T}B_T) + O_p(\delta) = O_p(\Delta_{\delta,T}B_T) = o_p(1)$$

as $\delta \to 0$ and $T \to \infty$, under Assumption 4.3.

Lastly, by Lemma A.2 (a) and boundedness of $K$, we have

$$\mathbb{E}[|R_{n,\delta}|] \leq \left( \sup_{s \in [-T,T]} \mathbb{E}[\Gamma_T(s)] \right) \int_{-\delta}^{-\delta-1} |K(s/B_T)| ds = O(\delta),$$

from which we deduce that $R_{n,\delta} = o_p(1)$. This completes the proof. \hfill \Box

**Proof of Proposition 4.2.** To show consistency of $\Lambda_{n,\delta}(r)$, it suffices to prove

$$\Lambda_T(r) \to_p \Lambda(r) \quad (53)$$

as $T \to \infty$ and

$$\Lambda_{n,\delta}(r) - \Lambda_T(r) = o_p(1) \quad (54)$$

as $\delta \to 0$ and $T \to 0$.

The proof of (53) is similar to that of Theorem 2.3 under small-$b$ asymptotics, because $\Lambda_T(r)$ is a nonparametric estimator of $\Lambda(r)$ using truncation kernel with bandwidth $A_T$ which follows small-$b$ asymptotics. Note the bias of $\Lambda_T(r)$ is

$$\mathbb{E}\Lambda_T(r) - \Lambda(r) = \int_{|s| \leq A_T} |s|^{r-1} \frac{T - |s|}{T} \Gamma(s) ds - \int_{-\infty}^{\infty} |s|^r \Gamma(s) ds = -(M_T + N_T), \quad (55)$$

where

$$M_T = \int_{|s| \leq A_T} \frac{|s|^{r+1}}{T} \Gamma(s) ds \quad \text{and} \quad N_T = \int_{|s| \geq A_T} |s|^r \Gamma(s) ds.$$

For $M_T$, we have

$$|M_T| \leq \frac{A_T}{T} \int_{|s| \leq A_T} |s|^r |\Gamma(s)| ds \leq \frac{A_T}{T} \int_{-\infty}^{\infty} |s|^r |\Gamma(s)| ds \to 0$$

as $T \to \infty$, since $A_T/T \to 0$. Moreover, we have $|N_T| \to 0$ as $T \to \infty$ since $A_T \to \infty$. Therefore, it follows from (55) that

$$|\mathbb{E}\Lambda_T(r) - \Lambda(r)| \to 0 \quad (56)$$

as $T \to \infty$. As for the variance of $\Lambda_T(r)$, we can show

$$\lim_{T \to \infty} \frac{T}{A_T} \text{var} (\Lambda_T(r)) \to 2 \left( \int_{-\infty}^{\infty} |s|^r \Gamma(s) ds \right)^2, \quad (57)$$

similarly as in the proof of Lemma 2.2 under small-$b$ asymptotics. Finally, (53) follows immediately from (56) and (57) since both bias and variance of $\Lambda_T(r)$ converges to zero as $T \to \infty$.

To show (54), we write

$$\Lambda_T(r) - \Lambda_{n,\delta}(r) = P_{n,\delta} + Q_{n,\delta} + R_{n,\delta},$$

where

$$P_{n,\delta} = \int_{|s| \leq A_T} |s|^{r-1} \frac{T - |s|}{T} \Gamma(s) ds,$$

$$Q_{n,\delta} = O_p(\Delta_{\delta,T}B_T) + O_p(\delta) = O_p(\Delta_{\delta,T}B_T) = o_p(1),$$

and

$$R_{n,\delta} = \int_{|s| \geq A_T} |s|^r \Gamma(s) ds.$$


where

\[ P_{n,\delta} = \delta^{1+r} \sum_{|j| \leq a_n} |j|^r |\Gamma_T(j\delta) - \Gamma_n(j)| \]

\[ Q_{n,\delta} = \sum_{j = -a_n}^{a_n-1} \int_{j\delta}^{(j+1)\delta} \|s|^r \Gamma_T(s) - |j\delta|^r |\Gamma_T(j\delta)| ds \]

\[ R_{n,\delta} = -\delta A_T \Gamma_T(A_T), \]

each of which will be analyzed in what follows. For \( P_{n,\delta} \), we have

\[ \mathbb{E} |P_{n,\delta}| \leq \left( \max_{|j| \leq a_n} \mathbb{E} |\Gamma_T(j\delta) - \Gamma_n(j)| \right) \delta^{1+r} \sum_{|j| \leq a_n} |j|^r = O(\Delta_{\delta,T})O(A_T^{1+r}) = O(\Delta_{\delta,T}T^{p(1+r)}), \]

by Lemma A.2 (c). Therefore \( P_{n,\delta} = o_p(1) \) under Assumption 4.4. Next, to analyze \( Q_{n,\delta} \), we write

\[ Q_{n,\delta} = F_{n,\delta} + G_{n,\delta}, \]

where

\[ F_{n,\delta} = \sum_{j = -a_n}^{a_n-1} \int_{j\delta}^{(j+1)\delta} |s|^r |\Gamma_T(s) - \Gamma_T(j\delta)| ds \]

\[ G_{n,\delta} = \sum_{j = -a_n}^{a_n-1} \int_{j\delta}^{(j+1)\delta} (|s|^r - |j\delta|^r) \Gamma_T(j\delta) ds. \]

First, we have

\[ \mathbb{E} |F_{n,\delta}| \leq \left( \max_{|j| \leq a_n} \sup_{s \in [j\delta,(j+1)\delta]} |\Gamma_T(s) - \Gamma_T(j\delta)| \right) \int_{|s| \leq A_T} |s|^r ds \]

\[ = O(\Delta_{\delta,T})O(A_T^{1+r}) = O(\Delta_{\delta,T}T^{p(1+r)}) \to 0 \]

as \( \delta \to 0 \) and \( T \to \infty \), by Lemma A.2 (b) and Assumption 4.4. Next, to analyze \( G_{n,\delta} \), we notice that for any \( |j| \leq a_n \) and \( |s - j\delta| \leq \delta \),

\[ |s|^r - |j\delta|^r = r|j\delta|^{r-1}(|s| - |j\delta|) + o(\delta A_T^{r-1}), \]

by first order Taylor expansion. Therefore, we have

\[ \max_{|j| \leq a_n} \sup_{s \in [j\delta,(j+1)\delta]} |s|^r - |j\delta|^r = O(\delta A_T^{r-1}), \]

from which and Lemma A.2 (a) we deduce that

\[ \mathbb{E} |G_{n,\delta}| \leq \left( \max_{|j| \leq a_n} \sup_{s \in (j-1)\delta,j\delta]} |s|^r - |j\delta|^r \right) \left( \max_{|j| \leq a_n} \mathbb{E} |\Gamma_T(j\delta)| \right) \int_{|s| \leq A_T} ds \]

\[ = O(\delta A_T^{r-1})O(1)O(A_T) = O(\delta T^{pr}) = o\left(\Delta_{\delta,T}T^{p(1+r)}\right) \to 0 \]

as \( \delta \to 0 \) and \( T \to \infty \), under Assumption 4.4. Therefore, it follows immediately from (58), (59) and
(60) that \( Q_{n, \delta} = o_p(1) \). Lastly, we deduce from Lemma A.2 (a) that

\[
\mathbb{E}|R_{n, \delta}| \leq \delta A_{\gamma} T \mathbb{E}[\Gamma_T(A_{\gamma})] = \delta A_{\gamma} T O(1) = O(\delta T^{\alpha\gamma}) = o\left(\Delta_{\delta, T} T^{\alpha(1+\gamma)}\right) \to 0
\]
as \( \delta \to 0 \) and \( T \to \infty \), under Assumption 4.4. Therefore, \( R_{n, \delta} = o_p(1) \). Consequently, we have shown \( A_T(r) - \Lambda_{n, \delta}(r) = o_p(1) \), as desired. This completes the proof. \( \Box \)

**Proof of Proposition 4.3.** To show consistency of \( \sigma_{n, \delta}^2 \), it suffices to show that \( \sigma_{n, \delta}^2 - \sigma_T^2 = o_p(1) \) since \( \sigma_T^2 \to_p \sigma^2 \) as \( T \to \infty \). Note that

\[
\sigma_T^2 = \frac{1}{T} \sum_{i=1}^{\infty} \int_{(i-1)\delta}^{i\delta} U_i^2 dt \quad \text{and} \quad \sigma_{n, \delta}^2 = \frac{1}{T} \sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} U_i^2 \Delta_{n, \delta}^2 dt - \frac{\delta}{T} U_0^2,
\]

from which it follows that

\[
\sigma_T^2 - \sigma_{n, \delta}^2 = \frac{1}{T} \sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} \left(U_i^2 - U_{n, \delta}^2\right) dt + \frac{\delta}{T} U_0^2.
\]

Moreover, we have

\[
\mathbb{E}\left[ \frac{1}{T} \sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} \left(U_i^2 - U_{n, \delta}^2\right) dt \right] \leq \frac{1}{T} \sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} \mathbb{E}\left|U_i^2 - U_{n, \delta}^2\right| dt
\]

\[
\leq \frac{1}{T} \sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} \mathbb{E}\left|U_i^2 + U_{n, \delta}\right| \left|U_i - U_{n, \delta}\right| dt
\]

\[
\leq \frac{1}{T} \sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} \mathbb{E}\left(U_i^2 + U_{n, \delta}\right)^2 \left[\mathbb{E}\left(U_i^2 + U_{n, \delta}\right)^2\right]^{1/2} dt
\]

\[
\leq \left[\sup_{\delta \geq 0} \mathbb{E}\left(U_i^2\right)^2\right]^{1/2} \left[\sup_{0 \leq s \leq T, t \leq \delta} \mathbb{E}\left(U_t - U_s\right)^2\right]^{1/2} \frac{1}{T} \int_{0}^{T} dt
\]

\[
= O\left(\Delta_{\delta, T}\right)
\]

and \( \delta U_0^2 / T = O_p(\delta/T) \). Therefore, we can deduce that

\[
\sigma_T^2 - \sigma_{n, \delta}^2 = O_p(\Delta_{\delta, T}) + O_p(\delta/T) = o_p(1)
\]
as \( \delta \to 0 \) and \( T \to \infty \), as desired.

Next, to establish consistency of \( \kappa_{n, \delta} \), it suffices to show \( \kappa_{n, \delta} - \kappa_T = o_p(1) \), since \( \kappa_T \to_p \kappa \) under Assumption 4.5 (a), as shown in Jeong and Park (2015). In particular, we will show

\[
\frac{1}{T} \sum_{i=1}^{n} \delta u_{i-1}^2 = \frac{1}{T} \int_{0}^{T} U_i^2 dt + o_p(1)
\]

(62)

where \( T^{-1} \int_{0}^{T} U_i^2 dt \to_p \sigma^2 > 0 \) as \( T \to \infty \), and

\[
\frac{1}{T} \sum_{i=1}^{n} u_{i-1}(u_i - u_{i-1}) = \frac{1}{T} \int_{0}^{T} U_i dU_i + o_p(1),
\]

(63)

42
from which we can deduce

\[
\kappa_{n, \delta} = \left( \sum_{i=1}^{n} \delta u_{i}^{2} \right)^{-1} \sum_{i=1}^{n} u_{i-1} (u_{i} - u_{i-1})
\]

\[
= \left( \frac{1}{T} \int_{0}^{T} U_{t}^{2} \, dt \right)^{-1} \frac{1}{T} \int_{0}^{T} U_{t} dU_{t} + o_{p}(1) = \kappa_{T} + o_{p}(1)
\]
as \delta \to 0 \text{ and } T \to \infty. \text{ First, we notice that}

\[
\frac{1}{T} \sum_{i=1}^{n} \delta u_{i-1}^{2} - \frac{1}{T} \int_{0}^{T} U_{t}^{2} \, dt = (\sigma_{n, \delta}^{2} - \sigma_{T}^{2}) - \delta \frac{U_{T}^{2}}{T} = o_{p}(1)
\]
by (61) and \( \delta U_{T}^{2}/T = O_{p}(\delta/T) = o_{p}(1) \), and then (62) follows directly. Secondly, to establish (63), we write

\[
\sum_{i=1}^{n} u_{i-1} (u_{i} - u_{i-1}) = \frac{1}{2} \left( u_{n}^{2} - u_{0}^{2} \right) - \frac{1}{2} \sum_{i=1}^{n} (u_{i} - u_{i-1})^{2}
\]

\[
= \frac{1}{2} \left[ (U_{T}^{2} - U_{0}^{2}) - \sum_{i=1}^{n} (U_{i\delta} - U_{(i-1)\delta})^{2} \right]
\]
(64)
and

\[
\int_{0}^{T} U_{t} dU_{t} = \frac{1}{2} (U_{T}^{2} - U_{0}^{2} - [U]_{T}),
\]
(65)
where \([U]\) denotes the quadratic variation of \(U\). By Ito formula, we have

\[(U_{i\delta} - U_{(i-1)\delta})^{2} = 2 \int_{(i-1)\delta}^{i\delta} (U_{t} - U_{(i-1)\delta}) dU_{t} + ([U]_{i\delta} - [U]_{(i-1)\delta}),\]

from which it follows that

\[
\sum_{i=1}^{n} (U_{i\delta} - U_{(i-1)\delta})^{2} - [U]_{T} = 2 \sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} (U_{t} - U_{(i-1)\delta}) dU_{t}.
\]

Moreover, under Assumption 4.5, we have, as shown in Chang and Park (2014), that

\[
\sum_{i=1}^{n} \int_{(i-1)\delta}^{i\delta} (U_{t} - U_{(i-1)\delta}) dU_{t} = O_{p}(pT \Delta_{\delta,T}) + O_{p}(\sqrt{qT \Delta_{\delta,T}}) + O_{p}(\sqrt{T})
\]
as \delta \to 0 \text{ and } T \to \infty. \text{ Consequently, we can deduce that}

\[
\frac{1}{T} \sum_{i=1}^{n} (U_{i\delta} - U_{(i-1)\delta})^{2} - [U]_{T} = O_{p}(pT \Delta_{\delta,T}) + O_{p}(\Delta_{\delta,T} \sqrt{qT/\Delta_{\delta,T}}) + O_{p}(1/\sqrt{T}) = o_{p}(1)
\]
(66)
as \delta \to 0 \text{ and } T \to \infty, \text{ by Assumption 4.5 (c). Finally, (63) follows from (64), (65) and (66). This completes the proof.}
C. Figures for Illustration and Simulation

Fig. 1. Discrete Time LRV Estimators for 1-Month Forward Premium of US/UK Exchange Rates and 3-Month US T-Bill Rates

Notes: Presented are discrete time LRV estimates $\Omega_n$ with bandwidth choices RT, NP and SP of 1-month forward premium of US/UK exchange rates from January 2, 1979 to June 30, 2015 (left panels) and 3-month US T-bill rates from January 4, 1954 to June 30, 2008 (right panels). The estimates are plotted against sampling intervals of the discrete samples used in the estimation. In particular, the sampling interval $\delta$ ranges from $1/252$ (daily frequency) to $1/2$ (semi-annual frequency). Parzen kernel is used.
Fig. 2. Simulated Means of Discrete Time LRV Estimators

Notes: Presented are simulated means of usual discrete time LRV estimators $\Omega_n$ using the discrete samples collected from the baseline OU (panel (a)) and SR (panel (b)) models. In both panels, results based on 1000 iterations are reported across sampling intervals ranging from $\delta = 1/252$ (daily frequency) to $\delta = 1/2$ (semi-annual frequency). In each of the 9 figures in $3 \times 3$ plots, results for three sampling spans ($T = 10, 30, 50$) and three bandwidth choices (RT, NP, SP) are presented respectively. Parzen kernel is used.
Fig. 3. Continuous Time LRV Estimates for 1-Month Forward Premium of US/UK Exchange Rates and 3-Month US T-Bill Rates

Notes: Presented are continuous time LRV estimates $\Pi_{n, \delta} = \delta \Omega_n$ with bandwidth choices CRT, CNP and CSP of 1-month forward premium of US/UK exchange rates from January 2, 1979 to June 30, 2015 (left panels) and 3-month US T-bill rates from January 4, 1954 to June 30, 2008 (right panels). The estimates are plotted against sampling intervals of the discrete samples used in the estimation. In particular, the sampling interval $\delta$ ranges from 1/252 (daily frequency) to 1/2 (semi-annual frequency). Parzen kernel is used.
Fig. 4. Simulated Biases of Continuous Time LRV Estimators

Notes: Presented are the simulated means of continuous time LRV estimators $\Pi_{n,\delta} = \delta \Omega_n$ with bandwidth choices CRT, CNP and CSP for the baseline OU (panel(a)) and SR (panel(b)) models. In each panel, results based on 1000 iterations across sampling intervals ranging from $\delta = 1/252$ (daily frequency) to $\delta = 1/2$ (semi-annual frequency) are presented in $3 \times 3$ plots for three sampling spans and three bandwidth choices. The true LRVs for the simulated models OU-B (with $\kappa = 4.8265$) and SR-B (with $\kappa = 0.1102$) are $1.5973 \times 10^{-6}$ and 0.014 respectively. Parzen kernel is used.
Fig. 5. Simulated Standard Errors of Continuous Time LRV Estimators

(a) OU Baseline Model

(b) SR Baseline Model

Notes: Presented are the simulated standard errors of continuous time LRV estimators $\Pi_{n,\delta} = \delta \Omega_n$ with bandwidth choices CRT, CNP and CSP for the baseline OU (panel(a)) and SR (panel(b)) models. In each panel, results based on 1000 iterations across sampling intervals ranging from $\delta = 1/252$ (daily frequency) to $\delta = 1/2$ (semi-annual frequency) are presented in $3 \times 3$ plots for three sampling spans and three bandwidth choices.
Fig. 6. Two-Week Local Means and Local Standard Errors of Continuous Time LRV Estimates for 1-Month Forward Premium of US/UK Exchange Rates and 3-Month US T-Bill Rates

Notes: Presented are two-week local means and local standard errors of continuous time LRV estimates \( \Pi_{n, \delta} \) using bandwidths CRT, CNP and CSP for 1-month forward premium of US/UK exchange rates (left panels) and 3-month US T-bill rates (right panels). For comparison, the scales of y-axes in panel (a) are set to be the same as those in Figure 3, which reports original LRV estimates of above two times series. Both local means (panel (a)) and local standard errors (panel (b)) are plotted.
Fig. 7. Simulated Means of Two-Week Local Standard Errors of Continuous Time LRV Estimators

Notes: Presented are simulated means of two-week local standard errors of continuous time LRV estimators $\Pi_{n, \delta} = \delta \Omega_{n}$ using bandwidth choices CRT, CNP and CSP for the baseline OU (panel(a)) and SR (panel(b)) models, which are based on 1000 iterations.